

UNIVERSIDADE FEDERAL FLUMINENSE
Instituto de Computação

Alexandre Costa Ferreira Lopes

**Um método para a geração de estimativas de reputação mais precisas perante
a oscilação de comportamento das entidades avaliadas**

Niterói
Julho de 2006

UNIVERSIDADE FEDERAL FLUMINENSE
Instituto de Computação

Alexandre Costa Ferreira Lopes

Um método para a geração de estimativas de reputação mais precisas perante a oscilação de comportamento das entidades avaliadas.

Dissertação de Mestrado submetida ao Programa de Pós-Graduação em Computação da Universidade Federal Fluminense como requisito parcial para obtenção do título de Mestre. Área de concentração: Otimização Combinatória e Inteligência Artificial.

Orientadora:

Prof.^a Ana Cristina Bicharra Garcia. Ph.D.

Niterói, RJ
Julho de 2006

Alexandre Costa Ferreira Lopes

Um método para a geração de estimativas de reputação mais precisas perante a oscilação de comportamento das entidades avaliadas.

Dissertação de Mestrado submetida ao Programa de Pós-Graduação em Computação da Universidade Federal Fluminense como requisito parcial para obtenção do título de Mestre. Área de concentração: Otimização Combinatória e Inteligência Artificial.

Banca Examinadora:

Prof^a. Ana Cristina Bicharra Garcia. Ph.D. (Orientadora)
Universidade Federal Fluminense

Prof^a. Claudia Lage Rebello da Motta. D.Sc.
Universidade Federal do Rio de Janeiro

Prof^a. Simone Diniz Junqueira Barbosa. D.Sc.
Pontifícia Universidade Católica – PUC-Rio

Niterói, RJ
Julho de 2006

Agradecimentos

À **Profª. Ana Cristina**, por suas orientações e pelo espírito constante de aperfeiçoamento e de esforço para a realização de um trabalho de qualidade.

Aos membros da banca examinadora, **Profª. Cláudia Motta e Simone Junqueira**, que gentilmente aceitaram colaborar com este trabalho.

Ao **Profº. Chrysanthos Dellarocas** que, mesmo à distância, muito ajudou com suas argumentações e com a “lapidação” de minhas idéias.

Ao **Profº. Carlos Lucena**, e à **Profª. Viviane Silva** que tão bem me acolheram na PUC, e ajudaram no meu desenvolvimento na área de agentes computacionais e de mercados eletrônicos.

Ao **Profº. Ferraz**, tanto como professor como colega de turma, que sempre contribuiu para o meu desenvolvimento acadêmico.

À minha querida esposa **Gabi**, meu “alicerce”, principalmente nos últimos anos de mestrado. Paciente e compreensiva, me influencia a cada dia com o exemplo de amor e de integridade que tem com nossa família e com o seu trabalho.

Aos meus pais **Waldyr e Florismar**, que tanto me amam. “Desde sempre” me ensinaram os valores básicos do homem, a importância do trabalho duro, da disciplina e da honestidade.

Ao meu amigo **Marcus Venícius**, pelas orientações e força dada ao longo do mestrado.

Aos meus amigos da UFF, **Leandro Neumann, José Luiz, Cristiano Maciel, Paula Yamada e Adriana Franco**, que me ajudaram a trilhar os bons caminhos do mestrado.

Aos meus amigos da Caixa Econômica Federal, em especial à **Regina Mehl, Mônica e Paulo Calçada, Cláudio Santos, Flávia Evaristo, Hugo Fernandes, Cristiane Martins, Marcelo Conceição e Márcio Nascimento**, cujo suporte permitiu conciliar minhas atribuições do emprego com as do mestrado.

Sumário

<i>Agradecimentos</i>	iv
<i>Sumário</i>	v
<i>Lista de Figuras</i>	vii
<i>Lista de Tabelas</i>	ix
<i>Resumo</i>	x
<i>Resumo</i>	x
<i>Abstract</i>	xi
Capítulo 1 Introdução	12
1.1 Filtragem Colaborativa Automática aplicada aos Sistemas de Reputação	13
1.2 O problema do “Falso Bom Vizinho” e a oscilação de comportamento	14
1.3 Objetivos, resultados esperados, e hipóteses de trabalho	16
1.4 Organização do trabalho	17
Capítulo 2 Conceitos Fundamentais	18
2.1 Sistemas de recomendação	18
2.1.1 Classificação de Sistemas de Recomendação	22
2.1.2 Técnica de Filtragem Colaborativa Automática	27
2.1.3 Vantagens e desvantagens da Filtragem Colaborativa	32
2.1.4 Novas pesquisas em Sistemas de Recomendação	34
2.2 Sistemas de Confiança e de Reputação	35
2.2.1 Conceito de Confiança	35
2.2.2 Conceito de Reputação	36
2.2.3 Sistemas de Reputação	37
2.2.3.1 Propriedades de Sistemas de Reputação	39
2.2.3.2 Classificação de Sistemas de Reputação	42
2.2.3.3 Desafios e áreas de pesquisas	45
2.3 Filtragem Colaborativa Automática aplicada aos Sistemas de Reputação	46
2.3.1 Estimativas de reputação personalizadas e Filtragem de Agrupamentos	48
2.3.2 Peso de Relevância (Significance Weighting)	50
2.3.3 Conclusões	51
Capítulo 3 Método Proposto	52
3.1 Modelo de alto nível do método proposto:	53
3.2 Transformação da Matriz de Avaliação e cálculo de reputação	54
3.2.1 Matriz de Avaliação estendida	54
3.2.2 Forma de realização da transformação	57
3.2.3 Exemplo de transformação da matriz de avaliação	58
3.3 Propriedades do método proposto	60
3.4 Escopo de atuação	61
Capítulo 4 Experimento	63
4.1 Organização da Simulação	63
4.2 Fases da Simulação	65
4.2.1 Fase de Negociação	65
4.2.2 Fase de Execução	65
4.2.3 Fase de Encerramento	67
4.2.3.1 Cálculo da Reputação do vendedor	67
4.2.3.2 Atualização dos recursos financeiros	68
4.3 Teste de geração de recomendações	69
4.3.1 Resultados	69
4.3.2 Conclusões	77
Capítulo 5 Trabalhos correlatos	78
5.1 Estimativas de reputação personalizadas e Filtragem de Agrupamentos	78

5.2	Peso de Relevância (Significance Weighting).....	79
5.3	Testes comparativos.....	79
5.3.1	Análises estatísticas	81
5.3.2	Teste comparativo 1.....	82
5.3.2.1	Resultados.....	83
5.3.2.2	Conclusões.....	92
5.3.3	Teste comparativo 2.....	93
5.3.3.1	Resultados.....	94
5.3.3.2	Conclusões.....	102
5.3.4	Conclusão dos testes comparativos	104
5.4	Outros trabalhos.....	104
5.4.1	Sistema ReGreT.....	104
5.4.1.1	Seleção de testemunhas	105
5.4.1.2	Credibilidade das testemunhas	106
5.4.1.3	Cálculo de agregação de reputação.....	107
5.5	Modelo de credibilidade de testemunhas.....	108
Capítulo 6	Conclusões.....	111
6.1	Contribuições.....	111
6.2	Limitações.....	112
6.3	Trabalhos futuros	113
Referências Bibliográficas.....		115
Anexo A	Modelagem do método proposto	119
	Diagrama de Classes.....	119
	Diagrama de Atividades	121
Anexo B	Modelagem da simulação	122
	Diagrama de Classes (Modelo de Análise).....	122
	Diagrama de Atividades	123
	Diagrama de Classes (Modelo de Design).....	125
	Classes do Sistema.....	129
	Diagramas de Sequência (Modelo de Design).....	130
	Interface de usuário do sistema.....	132
	Arquivo de parâmetros da simulação	132

Lista de Figuras

Figura 1.1 – Integração entre Sistemas de Reputação e Técnicas de Filtragem Colaborativa	16
Figura 3.1 – Modelo geral do método proposto	52
Figura 3.2 – Modelo conceitual completo (notação UML 2.0)	56
Figura 4.1 – Desempenho monetário dos compradores no cenário I	70
Figura 4.2 – Desempenho financeiro dos compradores no cenário II	70
Figura 4.3 – Desempenho financeiro dos compradores no cenário III	71
Figura 4.4 – Desempenho financeiro dos compradores no cenário IV	71
Figura 4.5 – Desempenho financeiro dos compradores no cenário V	72
Figura 4.6 – Desempenho financeiro dos compradores no cenário VI	72
Figura 4.7 – Quantidade de recomendações de vendedores Bad, Good e Neutral por rodada do cenário I	73
Figura 4.8 – Quantidade de recomendações de vendedores Bad, Good e Neutral por rodada do cenário II	74
Figura 4.9 – Quantidade de recomendações de vendedores Bad, Good e Neutral por rodada do cenário III	74
Figura 4.10 – Quantidade de recomendações de vendedores Bad, Good e Neutral por rodada do cenário IV	75
Figura 4.11 – Quantidade de recomendações de vendedores Bad, Good e Neutral por rodada do cenário V	75
Figura 4.12 – Quantidade de recomendações de vendedores Bad, Good e Neutral por rodada do cenário VI	76
Figura 4.13 – Queda do desempenho financeiro ao longo dos cenários de teste	77
Figura 5.1 – Comparação do desempenho financeiro entre os métodos testados no cenário I	85
Figura 5.2 – Comparação do desempenho financeiro entre os métodos testados no cenário II	86
Figura 5.3 – Comparação do desempenho financeiro entre os métodos testados no cenário III	87
Figura 5.4 – Comparação do desempenho financeiro entre os métodos testados no cenário IV	88
Figura 5.5 – Comparação do desempenho financeiro entre os métodos testados no cenário V	89
Figura 5.6 – Comparação do desempenho financeiro entre os métodos testados no cenário VI	90
Figura 5.7 – Distribuição de recomendações do método FCA2 no cenário II	91
Figura 5.8 – Distribuição de recomendações do método FCA2 no cenário III	91
Figura 5.9 – Comparação do desempenho financeiro dos métodos no cenário I	95
Figura 5.10 – Comparação do desempenho financeiro dos métodos no cenário II	96
Figura 5.11 – Comparação do desempenho financeiro dos métodos no cenário III	96

Figura 5.12 – Comparação do desempenho financeiro dos métodos no cenário IV	97
Figura 5.13 – Comparação do desempenho financeiro dos métodos no cenário V	97
Figura 5.14 – Comparação do desempenho financeiro dos métodos no cenário VI	98
Figura 5.15 – 1ª Rodada onde a quantidade de Bad é maior que a quantidade de Good + Neutral	98
Figura 5.16 – Distribuição de recomendações de FCA2 com priorização de Good e Neutral em relação a Bads	99
Figura 5.17 – Distribuição de recomendações de FCA1 com intercalação de Neutral e Bads nas recomendações intermediárias	100
Figura 5.18 – Distribuição de recomendações de FCA com intercalação de Neutral e Bads nas recomendações intermediárias	100
Figura 5.19 – Distribuição de recomendações de FCA3 com intercalação de Neutral e Bads nas recomendações intermediárias	101
Figura 5.20 – Menor quantidade de recomendações de vendedores Bad de FCA1 e FCA2 em relação a FCA e FCA3	101
Figura A.1 – Modelo conceitual completo (notação UML 2.0)	119
Figura A.2 – Fluxo de dados e de controle do método proposto (notação UML 2.0)	121
Figura B.1 – Visão conceitual da simulação (notação UML 2.0)	122
Figura B.2 – Fluxo de controle e de dados da simulação, com o Serviço de Reputação implementando o método proposto (notação UML 2.0)	123
Figura B.3 – Pacotes da simulação (notação UML 2.0)	125
Figura B.4 – Classe do pacote simulator.controller (notação UML 2.0)	125
Figura B.5 – Classes do pacote simulator.agents (notação UML 2.0)	126
Figura B.6 – Classes do pacote simulator.control (notação UML 2.0)	127
Figura B.7 – Classes do pacote simulator.util (notação UML 2.0)	128
Figura B.8 – Classes da simulação (notação UML 2.0)	129
Figura B.9 – Diagrama de seqüência da simulação (notação UML 2.0)	130
Figura B.10 – Diagrama de seqüência da transação entre comprador e vendedor (notação UML 2.0)	131

Lista de Tabelas

Tabela 1.1 – Matriz de avaliações de produtos em um sistema de recomendação	13
Tabela 1.2 – Preferências de avaliação de cada cliente	14
Tabela 1.3 – Exemplo de matriz de avaliação	15
Tabela 1.4 – Coeficiente de Similaridade entre o Cliente D e os demais (-1 menor similaridade e 1 maior similaridade)	15
Tabela 2.1 – Resumo das categorias de classificação	22
Tabela 2.2 – Matriz de avaliações de exemplo, considerando avaliações na escala de 1 a 7	29
Tabela 2.3 – Similaridade entre usuários	30
Tabela 2.4 – Matriz de avaliações de fornecedores em um sistema de reputação.	47
Tabela 3.1– Exemplos de Preferências de Reputação	55
Tabela 3.2 – Exemplo de Matriz de Avaliação Estendida	56
Tabela 3.3 – Exemplo de pesos atribuídos por um cliente a cada quesito de avaliação	58
Tabela 3.4 – Matriz de avaliação (resumida) do cliente ativo	58
Tabela 3.5 – Matriz de avaliação do cliente ativo 1 transformada	58
Tabela 3.6 – Correlação entre os clientes na matriz de avaliação original do cliente ativo 1	59
Tabela 3.7 – Correlação entre os clientes na matriz de avaliação ajustada do cliente ativo 1	59
Tabela 3.8 – Comparativo de similaridade de preferências com o Cliente 1	60
Tabela 4.1 – Forma de obtenção de avaliação para o item qualidade	67
Tabela 4.2 – Exemplo de obtenção da avaliação para o item preço	68
Tabela 4.3 – Descrição dos cenários de simulação	69
Tabela 5.1 – Resumo dos resultados dos testes	83
Tabela 5.2 – Média geral dos recursos financeiros por cenário	84
Tabela 5.3 – Resumo dos resultados dos testes	94
Tabela 5.4 – Quantidade de compradores com que cada vendedor interagiu durante as rodadas de treinamento	103
Tabela 5.5 – Quantidade de vendedores com que cada comprador interagiu durante as rodadas de treinamento	103

Resumo

A técnica de Filtragem Colaborativa Automatizada [1] vem sendo empregada com sucesso nos domínios de Filtragem de Conteúdo e de Recomendação de Produtos e Serviços, em sítios como Amazon [2] e iTunes [3]. Dellarocas [4] propõe a aplicação de filtragem colaborativa em sistemas de reputação para o cálculo de estimativas de reputação mais confiáveis e personalizáveis. Apesar de já solucionado no domínio de recomendação (com a atribuição de pesos aos coeficientes de similaridade [5]), o problema da determinação de avaliadores vizinhos pouco confiáveis (ou Falsos Vizinhos) encontra no domínio de reputação novas causas, como a variação de comportamento das entidades avaliadas. Ela pode tornar vizinhos distantes dois avaliadores com gostos completamente distintos e dois avaliadores com preferências bem distintas em vizinhos bem próximos, contribuindo para a diminuição da confiabilidade das estimativas de reputação.

Este trabalho apresenta um método baseado na técnica de filtragem colaborativa para tratar os problemas de falsa vizinhança, com as seguintes melhorias incorporadas: 1) inclusão de informações de gostos e preferências dos avaliadores no cálculo de reputação e na matriz de avaliações dos usuários, 2) realização de ajustes de vizinhança sobre a matriz de avaliações de acordo com a similaridade das preferências entre o usuário ativo e os avaliadores.

Experimentos são realizados através da simulação de um mercado eletrônico onde os compradores selecionam vendedores com base nas estimativas de reputação geradas pelo método proposto e por outros métodos correlatos. O objetivo principal do experimento é simular um ambiente de oscilação de comportamento dos vendedores, que permita comparar o desempenho do método proposto em relação às outras abordagens que também utilizam a técnica de filtragem colaborativa através da análise dos dados gerados pela simulação.

Abstract

The Automated Collaborative Filtering technique [1] has been successfully applied to Information Filtering and Recommendation domains, in web sites like Amazon [2] and iTunes [3]. Dellarocas proposes in [4] their use on reputation domains to provide more reliable and personalized reputation predictions. Despite being solved by researches of the recommendation field (through significance weighting [5]), the problem of selecting low-trusted (or false good/bad) neighborhoods finds new roots in the reputation domain, mostly related to the changing behaviour of the entities being evaluated. The changing behaviour can turn evaluators with similar tastes into distant ones, and evaluators with very distinct tastes into close neighbours, contributing to poor reputation estimates.

This work presents a method based on collaborative filtering to minimize those problems, with the following improvements: 1) information of raters' taste profiles are also used by the reputation prediction functions and added to the maintained user evaluation history; 2) execution of neighbourhood adjustments and transformations on the user evaluation history based on the similarities between the taste profiles of the active user and other evaluators.

Experiments are deployed through a simulated electronic market-place where buyers select sellers based on the reputation predictions generated by the proposed method and other correlated methods. The main goal of the experiment is to simulate the changing behaviour of sellers, and to compare in such environment the performance of the proposed method against other approaches that use collaborative filtering, through the analyses of data generated by the simulation.

Capítulo 1 Introdução

Sistemas de Reputação e de Confiança aplicados em sistemas de comércio eletrônico [4, 6-8] vêm sendo propostos com o objetivo de coibir a atuação de fornecedores com baixa qualidade de serviço prestado nos mercados eletrônicos. Esses sistemas podem combinar fontes de informação diretas e indiretas para melhor estimar a reputação dos fornecedores [9, 10]. As fontes diretas consistem em experiências obtidas em encontros passados entre o cliente e o fornecedor avaliado. As fontes indiretas consistem em informações obtidas através de testemunhos de outros clientes [4, 8, 11], ou da análise da rede de relacionamentos sociais mantida pelos fornecedores [9].

Apesar das fontes diretas serem consideradas a forma mais segura de determinar a confiabilidade de uma entidade, as informações disponíveis através das fontes indiretas existem em maior quantidade, devendo ser utilizadas em situações onde haja baixa probabilidade de quaisquer dois participantes terem um histórico de encontros passados [9].

Os modelos baseados em fontes indiretas são frágeis, dada a incerteza sobre a veracidade das informações obtidas, e pela dificuldade em se medir precisamente aspectos subjetivos como a qualidade dos serviços prestados. Muitas das abordagens estudadas trabalham com estimativas de reputação baseadas em agregações (somatórios ou médias) que não refletem as diferenças de gosto dos avaliadores ou o contexto das interações [4, 8]. Assuma por exemplo a seguinte situação: um casal em lua de mel se hospedou em um hotel 3 estrelas e acabou avaliando-o negativamente. Isto ocorreu porque o casal não conhecia o hotel e não se incomodava em pagar mais caro por um hotel com maior luxo. Agora imagine um grupo de mochileiros, que se importa mais com o preço do que com o luxo, tendo acesso à avaliação negativa do casal em relação ao hotel 3 estrelas. Nestes casos, as avaliações agregadas não seriam adequadas, fazendo com que os mochileiros desistissem de um hotel com boas chances de atender às suas necessidades.

Além de prejudicar as estimativas, as estimativas baseadas em agregações expõem fragilidades exploradas por usuários que se beneficiam de reputações acumuladas ou que se agrupam para exercer comportamento discriminatório perante outros usuários [4]. Se mal

calculada, a reputação pode fazer com que os clientes interajam com fornecedores que normalmente não escolheriam para realizar transações.

Para resolver estas fragilidades, alguns autores propõem incorporar técnicas de filtragem colaborativa aos sistemas de reputação, que propiciem a geração de estimativas de reputação personalizadas de acordo com o cliente.

1.1 Filtragem Colaborativa Automática aplicada aos Sistemas de Reputação

A filtragem colaborativa automática, denominada como FCA neste trabalho, vem sendo empregada em Sistemas de Recomendação e de Filtragem de Informações [1, 12]. Consiste em atribuir valores de avaliação personalizados para conteúdos e produtos com base na similaridade de perfis dos avaliadores. Identificam similaridades entre um usuário ativo, aquele que solicita a recomendação, e outros usuários – os avaliadores - com base na semelhança de avaliações passadas sobre itens em comum, e usam esta similaridade para gerar recomendações sobre itens ou conteúdos ainda não consumidos pelo usuário ativo. Os usuários vizinhos são aqueles que possuem maior grau de similaridade com o usuário ativo. A matriz de avaliação é o local onde são mantidas as avaliações dos usuários (Tabela 1.1).

Avaliador / Item avaliado	Usuário A	Usuário B	Usuário C
Produto 1	1	4	1
Produto 2	0	2	0

Tabela 1.1 – Matriz de avaliações de produtos em um sistema de recomendação

Na Tabela 1.1 os produtos 1 e 2 são avaliados pelos usuários A, B e C em uma escala de 0 a 4. É possível verificar que os usuários A e C emitem avaliações idênticas para os produtos 1 e 2, se tornando vizinhos. Com isto, se um deles avaliar bem um produto 3, o outro receberá uma recomendação deste produto. A técnica de FCA é descrita na seção 2.1.2 deste trabalho.

A FCA, ao ser incorporada em sistemas de reputação, visa estimar a reputação de fornecedores de forma personalizada, ou seja, calculando-a com base nas avaliações feitas sobre fornecedores por usuários com perfis similares aos do usuário ativo. A seção 2.3 deste trabalho

descreve em maiores detalhes a integração entre filtragem colaborativa e sistemas de reputação. Por hora, indicamos como exemplos desta integração os mecanismos de Reputação Colaborativa [13] e de Filtragem de Agrupamentos [4].

1.2 O problema do “Falso Bom Vizinho” e a oscilação de comportamento

Este problema é descrito como sendo a situação em que os avaliadores calculados como tendo maior similaridade com o usuário ativo, na realidade, não são tão similares [5, 14] . Esta situação pode ocorrer devido à coincidência de avaliações e à baixa quantidade de avaliações em comum entre os avaliadores. O próprio trabalho de Herlocker [5] já propõe soluções para esta questão. Entretanto, em sistemas de reputação existem outros fatores que podem contribuir para as situações de falsa vizinhança - como a oscilação de comportamento dos fornecedores - que os mecanismos de reputação que utilizam filtragem colaborativa precisam levar em consideração em suas abordagens.

Fornecedores que mudam de comportamento entre avaliações podem causar “ruídos” na identificação de similaridade entre clientes, ou seja, fazer clientes com preferências semelhantes não serem considerados vizinhos, ou de forma inversa, tornar vizinhos os clientes com preferências distintas. No contexto deste trabalho, o termo preferências representa o conjunto de quesitos (preço, qualidade, etc.) que foram utilizados pelo usuário para avaliar o desempenho da entidade avaliada. O exemplo a seguir ilustra esta situação de falsa vizinhança:

Na Tabela 1.2 as colunas preço e entrega representam os quesitos de avaliação de cada cliente, e as células representam o peso que os clientes dão a cada quesito. Observa-se que os clientes “D” e “E” possuem preferências semelhantes, e os Clientes “D” e “F” são pouco semelhantes quanto a suas preferências.

Cliente	Preço	Entrega
D	0,5	0,5
E	0,5	0,5
F	0,1	0,9

Tabela 1.2 – Preferências de avaliação de cada cliente

Com a observação da Tabela 1.3 verifica-se que “D” e “E”, mesmo tendo preferências similares, emitiram diferentes avaliações para um mesmo fornecedor “A”, não se tornando vizinhos. Os clientes “D” e “F”, mesmo com preferências bem distintas, emitiram avaliações equivalentes sobre o mesmo fornecedor “A”, se tornando vizinhos. Isto ocorreu porque o fornecedor “A” decidiu se comportar de forma diferente nos encontros com “D”, “E” e “F”, oferecendo serviço de baixa qualidade para “D” e “F”, e de boa qualidade para “E”.

	Cliente D	Cliente E	Cliente F
Fornecedor A	1	5	1
Fornecedor B	2	2	2
Fornecedor C	?	1	4

Tabela 1.3 – Exemplo de matriz de avaliação

Na prática “D” é muito mais similar a “E”, e a opinião de “E” deveria ter mais peso do que a de “F” na recomendação do Fornecedor “C”, entretanto, de acordo com a técnica de FCA (coeficiente de similaridade da Tabela 1.4), a opinião de “F” tem mais peso porque sua avaliação sobre “A” foi equivalente a de “D”, enquanto a avaliação de “E” não. Devido a esta distorção, nas próximas recomendações a chance de “D” ter uma decepção aumenta porque a tendência é de que “D” e “F” não tenham avaliações similares se os fornecedores avaliados mantiverem o mesmo comportamento entre as interações.

	Cliente E	Cliente F
Cliente D	-1	1

Tabela 1.4 – Coeficiente de Similaridade entre o Cliente D e os demais (-1 menor similaridade e 1 maior similaridade)

É importante observar que as situações observadas através deste exemplo não acontecem em Sistemas de Recomendação tradicionais porque os itens avaliados são produtos e não entidades racionais (humanas ou computacionais). Um produto não possui comportamento, suas características e aparência normalmente são mantidas a cada avaliação, enquanto um fornecedor é movido por objetivos que influenciam suas ações e comportamento [9]. Não estamos afirmando que a percepção do avaliador sobre um produto não mude ao longo do tempo, mas considerando um curto período de tempo, se dois clientes discordam na avaliação de um produto, é provável que o motivo da discordância seja o gosto diferenciado dos clientes. Já no caso da avaliação de

fornecedores não há garantias de que mantenham o mesmo comportamento perante cada avaliador, mesmo em curtos períodos de tempo.

Portanto, se o mecanismo de reputação, que incorporar a técnica de filtragem colaborativa, não levar em consideração os aspectos da oscilação de comportamento, poderá gerar estimativas de reputação imprecisas ou inadequadas que causem decepção aos clientes.

1.3 Objetivos, resultados esperados, e hipóteses de trabalho

Embora seja reconhecido na literatura que a incorporação de FCA aos Sistemas de Reputação é uma boa prática [4], o problema do “falso bom vizinho”, acentuado pelas mudanças de comportamento dos fornecedores, pode causar a diminuição da confiabilidade e precisão das estimativas de reputação.

Este trabalho propõe um método para minimizar este problema que consiste em ajustar a técnica de FCA para calcular de forma mais precisa a similaridade entre usuários, levando em consideração, além das avaliações emitidas por eles, as suas preferências representadas através dos quesitos de avaliação (preço, qualidade, etc.) que foram utilizados como base das avaliações emitidas, conforme ilustrado no esquema geral abaixo.

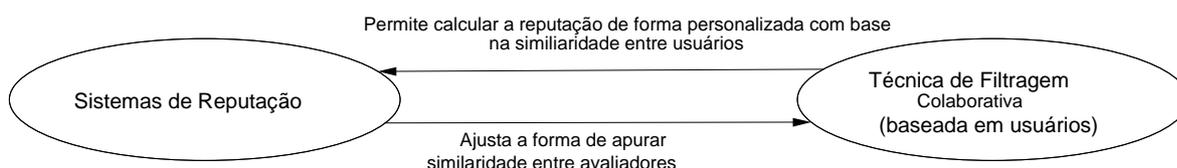


Figura 1.1 - Integração entre Sistemas de Reputação e Técnicas de Filtragem Colaborativa

Para verificar a efetividade do método proposto definimos as seguintes hipóteses centrais:

Hipótese 1 – O método proposto se comporta adequadamente em cenários de oscilação de comportamento por parte dos fornecedores, gerando estimativas de reputação adequadas aos gostos dos clientes.

Hipótese 2 – O desempenho do método proposto é melhor do que o dos métodos correlatos em cenários de oscilação de comportamento dos fornecedores.

Para testar estas hipóteses, foram realizados experimentos relatados nos capítulos 4 e 5, através da simulação de um mercado eletrônico, onde compradores selecionam vendedores que irão transacionar com base nas estimativas de reputação geradas por cada um dos principais mecanismos de reputação da literatura que utilizam filtragem colaborativa, descritos na seção 2.3 deste trabalho.

1.4 Organização do trabalho

Este trabalho está organizado da seguinte forma: no capítulo 2 são descritos em maiores detalhes os Sistemas de Recomendação e de Reputação, com ênfase para a descrição da técnica de filtragem colaborativa, e sua integração com sistemas de reputação. No capítulo 3 é apresentado o método proposto para minimizar os problemas de falsa vizinhança e oscilação de comportamento. No capítulo 4 são apresentados o experimento realizado sobre o método proposto e os resultados obtidos através da análise dos dados gerados. No capítulo 5 trabalhos correlatos são analisados em termos das semelhanças e diferenças para o método proposto e com base nos resultados obtidos em testes comparativos realizados entre eles. Também no capítulo 5 são apresentadas outras abordagens estudadas que não foram testadas nos experimentos. Por último, no capítulo 6 são apresentadas as conclusões, contribuições deste trabalho e necessidades de trabalhos futuros.

Capítulo 2 Conceitos Fundamentais

O avanço da Internet a torna cada vez mais presente na vida das pessoas. Seja para busca de conhecimento, compra de produtos, utilização de serviços, expansão de mercados, ou simplesmente para comunicação, cada dia mais pessoas e empresas recorrem à Internet para realizar estas tarefas. Se por um lado viabiliza novos modelos de negócio e propicia inúmeras oportunidades de desenvolvimento, a velocidade das transformações e o crescimento desordenado da Internet também cria o problema da sobrecarga de informações, que implica a dificuldade em se obter informação confiável, no tempo adequado e com o menor esforço possível.

A Internet é uma rede aberta onde milhões de pessoas de qualquer parte do mundo podem se conectar e desconectar a todo o momento. Estas pessoas, por muitas vezes, não se conhecem, e podem precisar estabelecer relações comerciais ou sociais, que muitas vezes possuem riscos inerentes, como por exemplo: acesso a informações incorretas, danos a imagem das pessoas, perda de dinheiro.

Neste contexto, Sistemas de Recomendação e de Reputação surgiram como uma forma de minimizar estes problemas, auxiliando aos usuários em suas escolhas: os Sistemas de Recomendação, por intermédio da filtragem de informações, e os Sistemas de Reputação, através do estabelecimento de medidas de confiabilidade. Neste capítulo introduzimos os conceitos básicos de sistemas de recomendação e de reputação, apresentamos uma revisão dos principais trabalhos conduzidos nesta área, perspectivas de estudo, e esforços de integração entre estes tipos de sistema.

2.1 Sistemas de recomendação

Os Sistemas de Recomendação (SR) visam gerar sugestões de itens de modo personalizado e de acordo com as preferências dos clientes. Os itens recomendados podem ser produtos, serviços ou conteúdos em geral. As recomendações têm um papel importante em situações onde os usuários por si só não possuam informações suficientes para a escolha ou aquisição de itens, e precisem

obter opiniões externas que auxiliem em suas decisões [15]. Surgiram como uma resposta ao problema da sobrecarga de informações.

A maioria dos Sistemas de Recomendação usa as opiniões fornecidas por especialistas ou outros usuários que tiveram experiências com os itens. Exemplos típicos de recomendações são as cotações de restaurantes e críticas de filmes, fornecidas por especialistas gastronômicos e críticos de cinema através de guias de restaurantes e de filmes. A propaganda boca a boca (*word-of-mouth*) também é um instrumento de recomendação, que é baseado nas opiniões de qualquer tipo de usuário (especialista ou não) sobre itens que teve contato.

Sistemas de Recomendação inicialmente foram desenvolvidos para os domínios de filtragem de notícias, de artigos e de conteúdos em geral. Suas primeiras versões eram referenciadas como sistemas de filtragem colaborativa [16], mas devido às estratégias e métodos empregados não se limitarem à filtragem colaborativa, passaram a ser denominados através de um termo mais amplo: Sistemas de Recomendação.

O primeiro SR desenvolvido foi o *Tapestry* [17]. Este sistema filtrava artigos distribuídos na Xerox PARC, com base em avaliações feitas pelos pesquisadores através de anotações. Possibilitava a geração de recomendações com base em consultas *SQL-like* que combinavam critérios como busca por palavra-chave e avaliações fornecidas por determinados funcionários. Por exemplo, seria possível fazer a seguinte consulta: “Quais livros o usuário X gostou que tenham a palavra Macintosh”. Entretanto a seleção de recomendadores ainda era feita de forma manual, ou seja, o sistema não era capaz de automaticamente identificar quais recomendadores eram mais similares ao usuário ativo (tinham opiniões mais similares), e por esta razão dependiam que os usuários informassem de quais recomendadores gostariam de obter opiniões. Esta característica particularmente não era um complicador, em vista da comunidade ser pequena, e todos os usuários se conheciam.

Posteriormente em 1994 foi desenvolvida a abordagem *GroupLens* [1], que recomendava artigos de um servidor de notícias com base nas avaliações feitas pelos seus usuários. Foi a

primeira abordagem a utilizar a técnica de FCA. A idéia básica aplicada era de que usuários que avaliassem itens de forma similar no passado, tenderiam a avaliar novos itens da mesma forma no futuro. Ao contrário do *Tapestry*, as avaliações não eram feitas através de anotações, e sim através de notas na escala de 1 a 5. Era capaz de automaticamente identificar as similaridades entre os usuários com base no histórico de avaliações feitas sobre artigos em comum, e a partir disto oferecer recomendações sobre artigos não lidos, mas altamente recomendados pelos usuários mais similares, denominados “vizinhos”.

Outra abordagem desenvolvida no mesmo período do *GroupLens* foi o sistema *Ringo* [18], que gerava, na Internet, recomendações personalizadas de álbuns musicais e artistas. As avaliações eram feitas na escala de 1 a 7, e o sistema também utilizava um algoritmo de filtragem de informações para determinar similaridades entre os usuários. Somente os usuários com coeficiente de similaridade maior do que um limiar definido eram selecionados para a geração das recomendações.

Posteriormente, começaram a ser desenvolvidos sistemas de recomendação para o domínio de comércio eletrônico na Internet. O desenvolvimento da *Web* mudou os modelos de negócio então existentes, com a transição do modelo de produção em massa, com produtos padronizados e mercado homogêneo, para um modelo de customização em massa, com maior variedade de produtos e opções para os clientes. Entretanto, esta evolução resultou na sobrecarga de ofertas e informações que os usuários eram expostos tornando o processo de seleção de itens mais complexo. O uso de SR neste contexto ajuda os clientes a encontrar, dentre uma grande variedade de produtos, aqueles com maior chance de satisfazer às suas necessidades [19].

Também se defendia a idéia de que Sistemas de Recomendação melhorariam as vendas não só com a facilitação da busca por produtos e serviços, mas também com o estabelecimento de vendas casadas, e com a fidelização dos clientes à medida que os sítios personalizem suas interfaces e recomendações de acordo com o perfil de cada usuário. Alguns exemplos de SR

neste domínio são os recursos “*Customers who Bought*” e “*Book Matcher*” da Amazon.com [20], e o recurso “*Feedback Profile*” do eBay [21].

O conceito de personalização ganhou força no ambiente da Internet, não só devido ao aumento de ofertas e da velocidade das informações proporcionado por este ambiente, mas também devido à flexibilidade e menor custo dos canais de acesso aos clientes (páginas e sítios *Web*). A personalização permitia que clientes encontrassem os produtos certos de forma mais rápida, deixando-os mais satisfeitos e aumentando a fidelização. A personalização envolve três etapas básicas [14]:

Identificar o cliente – Consiste em o sistema “perceber” quando um determinado cliente está presente. Dentre as formas utilizadas para detectar esta presença destacam-se o uso de *logins* e *cookies*, que apontam para informações sobre os perfis dos usuários conectados.

Descobrir os hábitos de consumo – Consiste em formar o perfil adequado de cada cliente. Isto é possível à medida que o cliente tenha interações com o sítio, e pode ser identificado de forma implícita ou explícita. Na forma implícita, o sistema automaticamente infere o perfil dos usuários com base no rastreamento dos seus hábitos de navegação ou no histórico de seleção de itens. Na forma explícita, o próprio usuário é responsável por informar ao sistema suas preferências. Alguns sítios como o *Amazon*, combinam as duas formas, estabelecendo o perfil inicial de forma explícita, e mantendo-o atualizado de forma implícita com o rastreamento das interações subsequentes entre o usuário e o sítio.

Gerar as recomendações – As recomendações podem ser geradas por solicitação explícita dos usuários ou à medida que selecionem determinados itens ou naveguem pelas áreas do sítio. A geração de recomendação consiste em efetuar o casamento entre o perfil do usuário identificado e as características dos itens disponíveis, e para isto são utilizadas técnicas de Inteligência Artificial como a Filtragem Colaborativa Automatizada (FCA), que será explicada mais à frente. Em alguns tipos de sistemas é necessário levar em conta os itens que o usuário já teve contato

anteriormente, e que não devem ser recomendados novamente (venda de filmes em DVD por exemplo).

2.1.1 Classificação de Sistemas de Recomendação

Diversas classificações foram propostas para Sistemas de Recomendação, sendo que nenhuma ainda foi aceita como padrão. Resnick [15] propôs uma classificação fundamentada nas principais características observadas nos Sistemas de Recomendação existentes na época (*Siteseer*, *GroupLens*). Dividiu as dimensões de classificação em Espaços de Domínio e de *Design Técnico*. As dimensões do espaço de domínio refletiam particularidades dos sistemas avaliados, como por exemplo: identificação das entidades que desempenhavam os papéis de recomendadores e clientes, os tipos de itens avaliados, a densidade e o tempo de vida (validade) das recomendações, etc. Já as dimensões do espaço técnico refletiam aspectos mais genéricos de SR, tais como a forma de agregação das avaliações, e a forma de uso e entrega das recomendações.

A classificação proposta em [22] e adaptada por [14], é considerada como uma das mais abrangentes, sendo definida a partir de uma taxonomia anteriormente proposta para sistemas de comércio eletrônico [19]. A Tabela 2.1 relaciona as categorias de classificação empregadas na taxonomia apresentada neste trabalho com as categorias do estudo prévio feito em [19].

Categoria de classificação	Denominação anterior feita para sítios de comércio [19]
Entrada/Saída	Interface de recomendação
Método de recomendação	Tecnologia de Recomendação
Grau de personalização	Tipos de recomendação
Entrega das recomendações	Encontrando recomendações Tipos de recomendação

Tabela 2.1 – Resumo das categorias de classificação

Categoria Entrada/Saída

Esta categoria classifica como as informações entram e saem do sistema de recomendação. Três subcategorias são consideradas: usuário ativo, comunidade de usuários, forma de oferecer recomendações.

- **Informações do Usuário ativo:** O usuário ativo é aquele que solicita as recomendações. Quanto maior for a quantidade de informações disponíveis sobre ele, maior será a chance de oferecer boas recomendações. Estas informações ajudarão o SR a determinar o perfil do usuário. Informações como sexo, idade e gostos podem ser coletadas pelos sistemas com esta finalidade.

Conforme já descrito anteriormente, existem duas formas de coleta de informações: implícita e explícita. Na forma implícita, o usuário não necessita exercer um papel ativo no fornecimento das informações, e muitas vezes não percebe que o sistema coleta automaticamente as informações. O Sistema obtém estas informações através da navegação do usuário (*links* visitados, tempo de permanência e navegação realizada dentro das páginas) ou através do histórico de compras e seleções de itens do sítio. Por exemplo: um usuário que solicite constantemente o aluguel de filmes de ação pode ser identificado como um aficionado no gênero de ação, ou o tempo que dedica para ler um determinado tipo de artigo pode ser considerado como interesse.

Na forma explícita, o usuário está ciente de que suas informações podem ser coletadas. Além de seus próprios dados, pode informar também avaliações sobre os itens que teve contato de modo que o sistema aprenda sobre a qualidade das recomendações fornecidas. Apesar de ser a forma mais segura de obtenção de preferências dos usuários, é a que requer mais esforço por parte dele.

- **Informações provenientes da comunidade de usuários:** As recomendações geradas pelo SR podem ser obtidas através de informações provenientes de toda a comunidade de usuários. Existem três tipos de informações: a popularidade externa de um item, comentários de texto e as avaliações.

A popularidade externa consiste em uma forma de recomendação não personalizada, uma vez que todos os usuários ativos recebem a mesma recomendação. É comum ser

empregada quando o SR ainda não conhece o perfil do usuário. Exemplos deste tipo são as listas de produtos mais vendidos.

Os comentários de texto consistem em observações solicitadas pelo SR que os usuários registram ao selecionar, adquirir ou conhecer um item. Trata-se de uma forma de obtenção de outras opiniões diferentes de especialistas ou editores de sítios eletrônicos. Por serem textuais, requerem um esforço maior do usuário ativo para julgar a opinião, e para o próprio avaliador ao registrar sua opinião.

As avaliações têm função similar aos comentários de texto, que é o registro da satisfação ou insatisfação do avaliador com o item avaliado. Entretanto, requerem menor esforço dos usuários por se tratarem de escalas numéricas de avaliação mais fáceis de serem informadas, e que podem ser agregadas por todos os usuários de forma a conferir uma avaliação geral ao item avaliado. As escalas podem variar de abordagem para abordagem: assumindo valores numéricos (-1, 0 e 1), valores categorizados (satisfeito, muito satisfeito, pouco satisfeito), binários (gostou ou não), e apresentando diferentes faixas (1 a 5, 1 a 7, etc.).

- **Forma de oferecer recomendações:** Existem duas formas ou modos de operação de recomendações: o modo de predição, onde o sistema estima um valor de avaliação para um determinado item, que corresponde ao valor provável que o usuário ativo daria ao avaliar o item; e o modo de recomendação, onde o sistema não recomenda um item, e sim uma lista de itens ordenados por aqueles melhor avaliados.

Duas variações do modo de recomendação são apresentadas em [19]. A primeira é *Top N List*, onde as listas podem ser mais personalizadas à medida que a quantidade máxima de itens seja determinada, dentre aqueles ainda não consumidos pelo usuário ativo, e que estes itens correspondam às preferências informadas. A segunda é *Ordered Search Results*, que consiste na geração de listas ordenadas de recomendações sem restrição da quantidade de itens, que permitam aos usuários visualizarem itens com menor chance de despertarem interesse. Entretanto a efetividade deste instrumento é questionável, uma vez que estudos

indicados em [23] mostram a grande maioria dos usuários apenas investigando as recomendações mais próximas do topo.

Categoria Método de Recomendação

Esta categoria classifica como as informações coletadas dos usuários e comunidade são usadas para gerar as recomendações. Os métodos principais métodos de recomendação são: seleção manual, resumos estatísticos, baseados em atributos, correlação item-a-item, e correlação usuário-a-usuário.

- **Seleção manual:** é o método utilizado por revisores, editores e especialistas para recomendar os produtos. Utilizam opiniões próprias e o conhecimento sobre o público-alvo para fazer as recomendações. Não é um método de recomendação personalizado.
- **Resumos estatísticos:** são médias ou totalizações das avaliações feitas sobre os itens disponibilizados no sítio suportado pelo SR. As avaliações utilizadas nos cálculos não se restringem àquelas feitas por especialistas e editores, podendo ser estendidas a todos os usuários do sistema.
- **Baseado em atributos:** é baseado nos atributos dos itens disponibilizados. Neste método os usuários podem informar os critérios de pesquisa para a recomendação ou o sistema pode recuperar o histórico de consumo dos usuários e inferir os critérios da recomendação. Por exemplo: um leitor assíduo de livros relacionados à Jornada nas Estrelas receberá recomendações de novas publicações relacionadas ao tema. O mesmo usuário também pode informar ao sistema que agora busca por jogos associados ao universo de Jornada nas Estrelas.
- **Correlação item-a-item:** visa identificar relações entre itens. Cada item consumido ou que o usuário ativo tenha demonstrado interesse pode servir de base para a recomendação de outros itens que estejam relacionados a ele. Dois itens podem estar relacionados porque um complementa o outro, ou porque são similares, por exemplo: um par de meias pode ser sugerido quando um tênis for adicionado ao cesto de compras, ou um livro sobre UML 2.0

pode ser sugerido quando um livro de UML tiver sido selecionado. No primeiro exemplo um produto complementa o outro, sendo geralmente adquirido em conjunto. No segundo caso existem vários livros similares sobre o mesmo assunto, porém não necessariamente iguais, por terem sido escritos por diferentes autores ou em épocas distintas.

- **Correlação usuário-a-usuário:** é uma forma de correlação que se baseia no princípio que se alguns usuários têm hábitos de consumo similares aos do usuário ativo, então é grande a chance de recomendarem novos itens que interessem ao usuário ativo. A técnica de Filtragem Colaborativa descrita mais a frente suporta este método de recomendação.

Categoria Grau de personalização

Esta categoria classifica o quão personalizadas são as recomendações geradas. Pode ser classificada em não-personalizada, efêmera e persistente. É intimamente ligada à categoria **Método de Recomendação**. Seus valores ajudam a representar aspectos específicos de cada método de recomendação. De acordo com o grau de personalização do SR, as recomendações geradas podem ser não-personalizadas, efêmeras ou persistentes.

As recomendações que são iguais para todos os usuários são consideradas como não-personalizadas. Os resumos estatísticos e a seleção manual apresentados na categoria Método de Recomendação são exemplos disto. As recomendações efêmeras são baseadas nas preferências do usuário no momento em que interage com o sistema, e sua geração não leva em consideração o histórico de interações do usuário. Os métodos de recomendação baseados em atributos e correlação item-a-item geram este tipo de recomendação. As recomendações persistentes são aquelas cuja geração se baseia no histórico de interações do usuário. São empregadas no método de recomendação correlação usuário-a-usuário.

A taxonomia originalmente proposta por [19] levava em consideração uma dimensão adicional, denominada automação, para classificar os métodos de recomendação sobre o ponto de vista dos usuários de SR. Nela, as recomendações podem ser automáticas ou manuais. As recomendações automáticas são aquelas em que o sistema requer esforço mínimo do usuário. As

manuais implicam em maior esforço do usuário para procurar recomendações. Os métodos de recomendação “Seleção Manual” e “Baseado em Atributos” são predominantemente classificados como manuais, os métodos “Resumos Estatísticos” e “Correlação usuário-a-usuário” são considerados automáticos, e o método Correlação item-a-item, dependendo do grau de esforço para seleção dos itens iniciais, pode ser classificado como manual ou automático.

Categoria Entrega das recomendações

As principais formas de entrega de informações são: *Pull*, *Push* e Orgânica. Na forma de recomendação *push* as recomendações são feitas aos usuários sem que eles tenham solicitado. Exemplos desta forma incluem as malas-diretas, janelas *pop-up* dos sites e as mensagens de correio eletrônico.

Na forma de recomendação *pull* os usuários solicitam as recomendações. Esta forma é considerada menos intrusiva do que a forma *push*, mas o usuário precisa de alguma forma ser avisado pelo sistema que existem recomendações disponíveis mediante sua solicitação.

Na forma orgânica, as recomendações são apresentadas sem maior alarde em conjunto com outros conteúdos do site. A medida que o usuário navegue normalmente pelo site, recomendações vão sendo geradas como parte do conteúdo do site. É uma das formas que requer o menor esforço do usuário, porém em algumas situações pode fazer com que a personalização do site não seja percebida pelo usuário.

À medida que o desempenho dos métodos de recomendação melhorou, os SR deixaram de usar exclusivamente a forma *pull*, e passaram a usar a forma *push*.

2.1.2 Técnica de Filtragem Colaborativa Automática

Conforme descrito no início deste capítulo, em uma situação real, as pessoas, por falta de conhecimento, procuram se aconselhar com amigos ou obter opiniões de especialistas para decidirem se devem consumir um determinado produto ou usufruir um serviço. O processo de pedir recomendações (*word-of-mouth* em inglês) suportado por computadores recebeu o nome de Filtragem Colaborativa. A filtragem ocorre através da seleção das recomendações mais

relevantes, aquelas oriundas de pessoas com maior similaridade com o usuário ativo. É colaborativa porque se apóia nas opiniões e avaliações de outras pessoas para selecionar os itens mais relevantes. É automática porque a similaridade entre pessoas é automaticamente calculada pelo sistema.

A FCA consiste em identificar similaridades entre um usuário ativo e outros usuários com base na semelhança de avaliações passadas sobre itens em comum, e usar esta similaridade para gerar recomendações sobre itens ainda não avaliados pelo usuário ativo. Para isto necessita:

- Armazenar avaliações sobre itens fornecidas pelos usuários do sistema em uma estrutura de dados denominada Matriz de Avaliações;
- Agrupar usuários com perfil semelhante (vizinhos), ou seja, que avaliam de forma semelhante os mesmos itens;
- Recomendar itens bem avaliados pelos vizinhos ao usuário ativo.

Existem vários algoritmos de implementação de FCA. Breese et al. [24] classifica vários algoritmos em duas classes gerais: baseados em memória, onde todo o histórico de avaliações é usado para gerar recomendações; baseados em modelos, onde o histórico de avaliações é usado para o aprendizado e construção de modelos que são usados para as recomendações. Como exemplos de algoritmos baseados em memória descreve, entre outros, o algoritmo de correlação, que é baseado na determinação de vizinhanças através da similaridade de usuários. Descreve também o algoritmo de similaridade de vetores, que é mais voltado para determinação de similaridade entre textos através de funções de co-seno. Como exemplos de algoritmos baseados em modelos descreve as Redes Bayesianas e os Modelos de Agrupamentos (*Cluster Models*).

Herlocker et al. [5] apresenta um *framework* de algoritmos para avaliação da precisão de diversas variações de algoritmos avaliados. Divide os algoritmos em componentes específicos como por exemplo: medição de similaridade (*Similarity Weighting*), peso de relevância (*Significance Weighting*), seleção de vizinhos (*Selecting Neighborhoods*), geração de recomendações (*Producing a Prediction*), que são avaliados quanto às suas variações

implementadas. Por exemplo: para medição de similaridade testa variações baseadas no Coeficiente de *Pearson* [25] e no Coeficiente de *Spearman*, para a seleção de vizinhos testa variações baseadas na seleção por número de vizinhos (*Best N-Neighborhoods*) e por limiar de similaridade (*Correlation-thresholding*).

Componente de medição de similaridade: consiste em avaliar a similaridade entre dois perfis de usuários. O perfil de cada um consiste no histórico de avaliações feitas sobre itens disponibilizados pelo sistema. Considerando a Tabela 2.2, o usuário Fernando tem como perfil as avaliações: 2, 1, e 6. Os produtos D e E não foram consumidos por ele e são passíveis de recomendação. Os produtos A e B não foram bem avaliados e o produto C foi bem avaliado.

	A	B	C	D	E
Fernando	2	1	6		
Luiz	2	2	5	5	2
Antônio	5	3	4	1	4

Tabela 2.2 – Matriz de avaliações de exemplo, considerando avaliações na escala de 1 a 7

Através da análise da matriz percebe-se maior similaridade de perfis entre Fernando e Luiz, de modo que a opinião de Luiz tende a ter maior influência nas recomendações sobre os itens ainda não consumidos por Fernando.

Uma das formas de medição de similaridade é através do Coeficiente de Pearson, que é um coeficiente que mede a força do relacionamento entre os perfis. A força destes perfis é representada em uma escala de -1 a 1, onde -1 representa forte correlação negativa, 0 a ausência de correlação e +1 forte correlação positiva. O cálculo do coeficiente é feito da seguinte forma:

$$(1) w_{a,u} = \frac{\sum_{i=1}^m (r_{a,i} - \bar{r}_a) * (r_{u,i} - \bar{r}_u)}{\sqrt{\sum_{i=1}^m (r_{a,i} - \bar{r}_a)^2} * \sqrt{\sum_{i=1}^m (r_{u,i} - \bar{r}_u)^2}}$$

Na equação (1), $w_{a,u}$ é o peso de similaridade calculado entre os perfis dos usuários a e u , $r_{a,i}$ e $r_{u,i}$ são os valores de avaliação que os usuários a e u respectivamente concederam ao item i , \bar{r}_a corresponde à média de avaliações do usuário a , e \bar{r}_u corresponde à média das avaliações do usuário u . Todas as médias devem ser consideradas apenas sobre os elementos que ambos os

usuários avaliaram. Por exemplo, em relação ao usuário ativo Fernando, a média do usuário Luiz é $(2+2+5)/3 = 3$, e a do usuário Antônio é $(5+3+4)/4 = 3$.

A correlação de Fernando para os demais é mostrada na Tabela 2.3:

Usuário	Correlação
Luiz	0.981981
Antônio	0.188982

Tabela 2.3 – Similaridade entre usuários

Uma outra forma, menos empregada, de medir similaridade é através da função de cosseno. Esta função consiste em calcular o ângulo entre dois vetores ou retas, que consistem no histórico de avaliações de cada usuário. Retorna valores na escala de 0 a 1, onde 0 determina menor correlação e 1 máxima correlação.

Componente de seleção de vizinhos: consiste em selecionar o conjunto de avaliadores ideais para geração da recomendação. Com base na similaridade calculada para cada avaliador, utiliza um dos dois métodos abaixo descritos para agrupamento de vizinhança:

- **Limiar de Similaridade (*Correlation-thresholding*):** define-se um limiar de similaridade que irá determinar quais avaliadores serão incluídos na vizinhança do usuário ativo. No exemplo acima se o limiar fosse definido como 0.50, Luiz (0.98) seria incluído na vizinhança de Fernando, mas Antônio (0.19) não. A vantagem desta abordagem é a garantia de seleção avaliadores realmente similares na vizinhança, o que aumenta a chance de realizar boas recomendações. O problema é a possibilidade do limiar definido ser restritivo e o conjunto de vizinhos ser muito pequeno ou até vazio.
- **Número de vizinhos (*Best N-Neighborhoods*):** consiste na seleção dos melhores avaliadores independente do limiar de correlação, apenas limitados a uma quantidade máxima definida, que idealmente deve variar entre 20 e 50 dependendo do domínio [5]. A ideia por trás desta abordagem é garantir um conjunto mínimo de avaliadores, entretanto, nem sempre será possível garantir que este conjunto seja apenas composto por avaliadores similares ao usuário ativo. Para efeito

ilustrativo, no exemplo acima, se o número definido fosse 2, Antônio, mesmo tendo baixa correlação (0.19), seria incluído na vizinhança de Fernando.

Componente de geração de recomendações: conforme apresentado anteriormente na classificação de sistemas de recomendação, uma das formas de classificar SR é através da forma de produção das recomendações. As principais formas são: predição, onde o sistema estima um valor de avaliação para um determinado item que corresponde ao valor provável que o usuário ativo daria ao avaliá-lo; e recomendação, onde o sistema não recomenda um item, e sim uma lista de itens ordenados por aqueles melhor avaliados. Por exemplo: no modo preditivo o usuário Fernando escolheria o item D ou E, e seria informado sobre o valor preditivo do item escolhido, no modo de recomendação receberia uma lista ordenada decrescentemente pelos valores preditivos apurados para D e E.

No modo de predição o cálculo é feito para o item selecionado através da média ponderada das avaliações dos vizinhos, com a seguinte equação:

$$(2) p_{a,i} = \bar{r}_a + \frac{\sum_{u=1}^m (r_{u,i} - \bar{r}_u) * w_{a,u}}{\sum_{u=1}^m |w_{a,u}|}$$

Na equação (2), $p_{a,i}$ corresponde ao valor preditivo do item i para o usuário ativo, m corresponde ao número de vizinhos, \bar{r}_a é a média das avaliações do usuário ativo, $r_{u,i}$ é a avaliação do vizinho sobre o item, \bar{r}_u é a média das avaliações do vizinho e $w_{a,u}$ é o fator de similaridade entre o vizinho e o usuário ativo (por Coeficiente de Pearson ou Co-seno).

Da mesma forma que no cálculo de similaridade, todas as médias devem ser consideradas apenas sobre os elementos que ambos os usuários avaliaram. As avaliações dos vizinhos são subtraídas das médias de modo a padronizar os valores de avaliação e evitar que diferentes distribuições de avaliações dos vizinhos causem distorção no valor preditivo final. Por exemplo: um vizinho pode ser mais exigente e somente submeter valores de avaliação em uma escala entre 1 e 3, enquanto outro pode ser menos exigente e submeter avaliações entre 3 e 5. Para garantir

que o valor preditivo agregado esteja na mesma escala dos valores de avaliação, a média ponderada é somada à média das avaliações do usuário ativo.

No modo de recomendação o cálculo acima descrito precisa ser realizado para cada um dos itens que o usuário ativo ainda não consumiu ou utilizou, e posteriormente é gerada uma lista ordenada pelos valores preditivos de cada item para o usuário ativo.

2.1.3 Vantagens e desvantagens da Filtragem Colaborativa

As vantagens provenientes da geração de recomendações personalizadas já foram citadas ao longo deste capítulo. Algumas vantagens adicionais podem ser percebidas, como a geração de recomendações independentes do conteúdo dos itens avaliados (texto, imagem, som), e a possibilidade de gerações de recomendações inesperadas, porém apreciadas pelo usuário ativo. Isto pode acontecer porque o sistema ao identificar perfis de consumo similares, abre a possibilidade de cada usuário similar recomendar itens que o outro usuário normalmente não buscaria, por não fazer parte de suas preferências registradas.

Dentro do conjunto de problemas inerentes à técnica de FCA, destacamos a esparsidade de informações de avaliação, a primeira avaliação (*cold start*), preferências de consumo efêmeras, o tempo de geração de recomendações.

Esparsidade de informações de avaliação: esparsidade é expressa pelo índice de ocupação da matriz de avaliação. Quanto menor for a esparsidade, maior será a quantidade de itens em comum consumidos ou utilizados pelos usuários. Uma matriz com 20% de esparsidade significa que de um universo de todas as possíveis ações de consumo pelos usuários, 80% ocorreram e resultaram em avaliações. Em razão da esparsidade se torna complicado recomendar itens e identificar vizinhanças. Se um produto não foi consumido por nenhum usuário, não pode ser recomendado. Por outro lado, quanto menor for a quantidade de avaliações menor será a chance de haver itens avaliados em comum, e conseqüentemente mais difícil será encontrar vizinhos dos usuários ativos.

Primeira avaliação: da mesma forma que um item não consumido não pode ser recomendado por vizinho algum (na técnica de FCA), se um usuário não consumir item algum, não será possível comparar os seus hábitos de consumo com outros usuários inviabilizando a determinação de vizinhanças. Sem vizinhança estabelecida não há como gerar recomendações personalizadas através de FCA.

Soluções alternativas para estes dois primeiros problemas envolvem a aplicação de recomendações não-personalizadas (como sumários estatísticos, seleção manual, etc.), ou baseadas em preferências informadas de forma explícita pelo usuário, até que haja informações suficientes no sistema para geração de recomendações através da FCA.

Preferências de consumo efêmeras: existem situações onde um usuário deseja selecionar um item pensando em um amigo ou parente que deseje presentear. Nestes casos as preferências que guiam o processo de seleção podem ser momentaneamente distintas às preferências habituais do usuário ativo. Solução para este problema envolve o sistema automaticamente perceber ou explicitamente ser informado sobre a intenção do usuário naquele momento, e passar a gerar recomendações com base nos métodos de recomendações baseados em atributos e correlação item-a-item, que são mais adequados para este cenário.

Tempo de geração de recomendações: a medida que as recomendações são geradas a cada pedido do usuário, e devido ao grande volume de itens disponibilizados e usuários registrados em SR¹, a FCA requer uma capacidade computacional muito elevada para que as recomendações sejam geradas dentro de intervalos de tempo aceitáveis. Soluções para este problema envolvem o emprego de algoritmos alternativos baseados em modelos [16, 24], que através de modelos de dimensões reduzidas, inferidos e atualizados com base nas avaliações, passem a gerar recomendações de forma mais rápida.

¹ No sítio da Amazon por exemplo existem mais de 3 milhões de produtos disponibilizados, e uma base de clientes cumulativa superior a 30 milhões.

2.1.4 Novas pesquisas em Sistemas de Recomendação

Grande foco vem sendo dado às pesquisas sobre métodos de recomendação, confiabilidade e precisão de recomendações. Novos campos de pesquisa vêm sendo explorados, como por exemplo a explanação das recomendações, que implica a mudança da abordagem *black-box* dos métodos atuais para uma abordagem que promova a transparência das interfaces de recomendações [12], com o fornecimento, complementar à recomendação, de informações sobre a confiabilidade dos recomendadores, entre outras. Este mesmo trabalho cita estudos que indicam aumento do índice de satisfação e de confiança dos usuários do SR com a aplicação destas abordagens.

Campos de pesquisa já estudados como a Esparsidade e Desempenho das Recomendações ganham novas abordagens, com a definição de algoritmos (*spreading-activation*) [16] que conseguem usar avaliações de itens similares, porém distintos, para geração de recomendações personalizadas em ambientes com esparsidade de informações.

Outro assunto que vem sendo pesquisado é a atualização de perfis ao longo do tempo [14]. As pesquisas giram em torno de questões sobre como acompanhar e rapidamente se adaptar às evoluções e mudanças de perfis dos usuários de modo que as recomendações geradas sempre fiquem próximas às suas necessidades atuais dos usuários.

A confiabilidade das recomendações também é um assunto bastante pesquisado em função da evolução dos modelos e sistemas de confiança e reputação. A integração das abordagens de recomendação e de reputação/confiança é tema de vários trabalhos [4, 12, 13], dada as ameaças que os mecanismos de reputação e de recomendação passam a sofrer em ambientes abertos com a presença de usuários com comportamentos desonestos e mal intencionados. Alguns pesquisadores [12] vêem os mecanismos de reputação como forma de garantir confiabilidade e veracidade das avaliações feitas sobre os itens em SR, e conseqüentemente de garantir a geração de recomendações de qualidade para os usuários ativos. Outras linhas de pesquisa [4, 13] destacam os benefícios que os SR proporcionam aos

mecanismos de reputação, principalmente com relação à geração de estimativas de reputações mais precisas e confiáveis, as quais são objetos de estudo detalhado neste trabalho.

2.2 Sistemas de Confiança e de Reputação

O provimento de serviços on-line via Internet e os mercados on-line formados através de sistemas de transações eletrônicas promovem encontros entre clientes e fornecedores remotamente dispostos e que muitas vezes nunca se encontraram antes [13, 26]. Os clientes geralmente não têm informações suficientes sobre os fornecedores, nem têm a possibilidade de verificar a qualidade dos serviços e produtos oferecidos antes de pagar por eles – somente após a transação saberão se o produto oferecido confere com a descrição, se está em bom estado, ou se foi entregue dentro do prazo acordado. Em função disto, estas interações incorrem em riscos [8], geralmente assumidos pelos clientes, uma vez que os fornecedores sabem que receberão pagamentos em dinheiro em troca do provimento dos serviços, resultando em uma assimetria de informações [26].

Sistemas de Confiança (*Trust*) e de Reputação ajudam a minimizar ou eliminar estes problemas. Ajudam os clientes a decidir se podem confiar nos fornecedores mesmo sem conhecê-los e sem ser possível experimentar antecipadamente o produto. Adicionalmente estes sistemas encorajam o bom comportamento dos fornecedores, uma vez que “a expectativa de que os clientes considerem o comportamento passado dos fornecedores em futuras interações restringe o comportamento atual dos fornecedores”[8].

2.2.1 Conceito de Confiança

A Confiança define, sob o ponto de vista de um indivíduo, o quanto ele confia em outro indivíduo. É uma crença que um determinado indivíduo tem sobre a atitude positiva (benevolente e honesta) do outro indivíduo em relação aos dois. Representa a probabilidade subjetiva de que um indivíduo A espera que um indivíduo B se comporte de forma a contribuir para o bem estar de A [26].

A Confiança é a consequência da confiabilidade de um indivíduo ou entidade. A confiabilidade se traduz pela capacidade, ou qualidade, de uma entidade ser confiável, ou seja, de ter as atitudes positivas perante às entidades que dela dependem.

O trabalho em [26] introduz uma nova noção de Confiança, denominada Confiança de Decisão (*Decision Trust*), inspirada no trabalho conduzido em [27], que consiste em definir o quanto uma entidade está propensa a confiar em outra, mesmo sabendo dos riscos disto. Insere as noções de utilidade (se vale a pena correr os riscos), e de atitude perante o risco para as entidades que confiam nas outras.

2.2.2 Conceito de Reputação

A Reputação corresponde ao que é geralmente dito ou sabido sobre o caráter ou posição de um indivíduo ou entidade. A Reputação pode ser considerada como uma medida coletiva (senso comum) da confiabilidade de uma entidade, baseada em avaliações feitas por membros da comunidade em que se situa [26].

A diferença entre Confiança e Reputação fica mais nítida a partir de agora. Enquanto a Confiança é um julgamento pessoal e subjetivo de cada entidade, a reputação reflete a visão da comunidade sobre o indivíduo, que nem sempre é equivalente à Confiança depositada por alguns indivíduos, conforme ilustrado na sentença (2) abaixo:

- (1) “eu confio em você por sua boa reputação”
- (2) “eu confio em você apesar de sua má reputação”

As sentenças acima também reforçam as afirmações feitas na introdução deste trabalho: que a Confiança pode ser medida através de várias fontes e evidências, diretas ou indiretas, públicas ou particulares. Na sentença (1) a entidade baseia a Confiança exclusivamente em informações de reputação: públicas porque são de conhecimento da comunidade e indiretas porque refletem a opinião de terceiros; enquanto na sentença (2), provavelmente a entidade se baseia em conhecimentos particulares ou experiências diretas que têm maior peso em relação à reputação aferida no ambiente.

Apesar de uma forma geral as fontes de informações diretas (baseadas em experiências pessoais) serem mais valorizadas do que as fontes indiretas, existem situações onde a escassez ou inexistência de experiências pessoais pode fazer com que as fontes indiretas sejam a base para a aferição de Confiança.

Existem abordagens que incorporam mecanismos de reputação em sistemas de confiança como o ReGreT [28], e vice-versa como o componente de Reputação por Testemunho implementado no Modelo FIRE [11]. Estas abordagens serão explicadas em detalhes no capítulo “Trabalhos Correlatos”.

2.2.3 Sistemas de Reputação

Sistemas de Reputação coletam, distribuem e agregam avaliações sobre o comportamento passado de entidades no contexto das interações ocorridas no ambiente monitorado. Desta forma, ajudam os usuários a avaliar a confiança que depositam em outros participantes, e a decidir sobre a efetivação de interações em cenários de incertezas e riscos. A seguir são apresentados alguns sistemas comerciais que utilizam tecnologias de reputação.

Fórum de *feedback* do eBay: *eBay* [21] é um sítio de leilões bastante popular que promove transações entre compradores e vendedores. O fórum de *feedback* permite que as partes envolvidas na transação consultem a reputação de cada um antes de iniciá-la, e ao seu término avaliem o desempenho de cada um como positiva, negativa ou neutra. A reputação de cada participante é obtida com a subtração do total de avaliações positivas pelo total das avaliações negativas. Além disto cada participante pode deixar comentários a respeito da outra parte envolvida na transação, que combinados com a avaliação podem oferecer mais subsídios para os usuários em futuras transações. O total de avaliações positivas, negativas e neutras também é apresentado em várias opções de janelas de tempo, como por exemplo: os últimos seis meses, último mês, e última semana.

Sítios de Revisão de Produtos: Sítios como *Epinions* [29] e *Amazon* [20] disponibilizam um conjunto de revisores que ajudam os clientes a adquirir produtos. Com base nestas

recomendações os revisores são avaliados pelos clientes, assim como os produtos que recomendam.

Epinions: é um sítio que disponibiliza revisões de produtos e lojas. Os fornecedores são cobrados pela quantidade de vezes que os usuários visitam suas páginas em busca de informações sobre os produtos fornecidos. As revisões de produtos e de lojas são feitas por membros que desejem assumir esta responsabilidade. As revisões consistem em comentários textuais e avaliações na escala entre 1 e 5 estrelas para cada uma das categorias de avaliação, por exemplo: para produtos podem ser usadas categorias como facilidade de uso, tempo de vida de bateria, etc. Para lojas podem ser avaliados aspectos referentes ao serviço oferecido, como entrega no prazo, facilidade de fazer pedido, etc.

Estas revisões por sua vez também são passíveis de avaliação por outros membros. Com base na utilidade das revisões, o revisor pode obter diferentes status, como Consultor, Revisor Principal, ou Principal da Categoria (melhor categoria). A seleção de revisores para cada usuário pode ser gerenciada através do recurso de Rede de Confiança, onde cada membro pode informar o grau de confiança depositado em cada revisor, e até mesmo bloqueá-los. Isto também influencia no status dos revisores descrito anteriormente, uma vez que quanto maior for a quantidade de membros que confiam em um revisor, melhor será seu status.

Existe também um mecanismo de incentivos diferenciado para o fornecimento de revisões honestas (negativas ou positivas). Diferenciado porque resulta em remuneração monetária ao revisor de acordo com a quantidade de usuários que são ajudados a encontrar produtos, ou a evitar comprar certos produtos. Este mecanismo funciona de forma automática, com base no uso que os usuários solicitantes fazem das recomendações geradas.

Sítio da Amazon: é essencialmente uma livraria on-line onde revisores podem disponibilizar suas revisões sobre livros vendidos no sítio. Revisões consistem em comentários em texto e avaliações na escala de 1 a 5 estrelas. Qualquer membro pode se registrar como revisor. A avaliação dos livros é obtida através da média de avaliações.

Da mesma forma que no *epinions*, as revisões podem ser julgadas com base em sua utilidade, inclusive por usuários que não sejam registrados como membros do sistema. Para cada revisão são apresentados os totais de votos e os totais de avaliações que consideraram a revisão como útil. Cada revisor pode ser classificado entre os 1000 melhores, 500 melhores, 100 melhores, 50 melhores, 10 melhores ou primeiro, com base na quantidade de avaliações que consideraram suas revisões como úteis. De forma similar às redes de confiança do *epinions*, existe um serviço que permite a cada usuário indicar seus revisores favoritos, o que também contribui na classificação dos revisores, entretanto, este mecanismo não utiliza incentivos financeiros para os revisores.

É importante salientar que a semelhança de atuação entre Sistemas de Recomendação e de Reputação - ambos procuram garantir aos clientes o fornecimento de serviços de qualidade - faz com que alguns destes sistemas sejam classificados tanto como Sistemas de Reputação quanto Sistemas de Recomendação. Como exemplo disto citamos os trabalhos de [19] e [26], que classificam de forma diferente o recurso de *feedback-profiling* do sítio *eBay*, o primeiro como Sistema de Recomendação e o segundo como Sistema de Reputação.

2.2.3.1 Propriedades de Sistemas de Reputação

Segundo Resnick [8], para funcionarem de modo adequado, os sistemas de reputação necessitam de pelo menos três propriedades abaixo descritas:

- **Longevidade:** Representa a presença permanente das entidades que possibilite formar uma expectativa de futuras interações. Esta propriedade se reflete na impossibilidade ou dificuldade de apagar o comportamento passado através de mudanças de identidades e de pseudônimo. Em muitos sítios os usuários se registram várias vezes, informando diferentes pseudônimos em cada registro, de modo a eliminar seu histórico de mau comportamento [8]. Soluções para estes problemas incluem: utilização de nomes reais, impossibilitar a troca de nomes e o uso de múltiplos pseudônimos, estabelecimento de taxas de adesão para novos

usuários, vinculação do ganho de reputação de novos usuários à aceitação de maior risco ou de preços mais caros.

- **Coleta e distribuição de *feedback* das interações correntes:** esta propriedade depende da vontade dos participantes em fornecer as avaliações após a realização das transações, o que implica a necessidade de estabelecimento de incentivos, como créditos em futuras transações, etc. Além disto, em muitos casos é difícil haver avaliações negativas, por medo de retaliação ou de indisposição com o participante avaliado. Soluções para estas situações foram propostas como o Controle de Anonimato [4].

A distribuição de *feedback* deixa de ser limitada com o aumento da integração dos sistemas de reputação, que permitem maior disseminação e unificação das informações de reputação dos usuários.

A coleta e a distribuição não costumam ser problemas em ambientes de reputação centralizados, mas ainda são aspectos complexos de serem viabilizados em ambientes descentralizados (redes ponto-a-ponto e sistemas multiagentes), em função de não haver protocolos padronizados para o fornecimento de *feedback* [26].

- **Agregação por *feedback*:** Estes mecanismos de computação de reputação são fundamentais para guiar decisões de confiança dos usuários. A seção “Classificação de Sistemas de Reputação” deste capítulo descreve em maiores detalhes os tipos de agregação de reputação, porém, neste momento, é interessante ressaltar que o desafio consiste em fazer com que estas medidas de reputação sejam úteis para os usuários decidirem em quem confiar. Para tal, questões como a precisão das medidas [28], robustez contra manipulação de avaliações e avaliadores desonestos [4], aferição de reputação dos avaliadores [13], recentidade das avaliações [11], são endereçadas por vários trabalhos de modo a tornar estes mecanismos mais efetivos e robustos.

O estudo em [30] define critérios de julgamento da qualidade e solidez dos sistemas de reputação:

- **Precisão no desempenho a longo prazo:** o histórico comportamental dos membros deve ser levado em conta na computação de reputação, de modo que o sistema seja capaz de diferenciar um novo membro da comunidade de um membro com histórico de baixo desempenho.
- **Recentidade:** é a capacidade de refletir o comportamento atual dos membros na computação de suas reputações. Reconhecer tendências atuais de comportamento e conferir maior peso a elas do que ao histórico já datado. Evitar que comportamentos mais antigos façam com que a qualidade das estimativas de reputação diminua e não corresponda ao comportamento real observado.
- **Suavidade (*Smoothness*):** garantir que variações momentâneas de comportamento não mudem significativamente a reputação do membro avaliado, até o momento em que se confirmem como tendências.
- **Robustez contra ataques:** é a capacidade dos sistemas resistirem a ataques de indivíduos ou grupos que exploram vulnerabilidades do sistema. Os autores afirmam que dentre os critérios acima citados, a Robustez Contra Ataques é o que requer esforço permanente em vista da contínua evolução das técnicas de ataque. Exemplos destes comportamentos são [4]:

avaliações desonestas positivas (*ballot stuffing*) : fornecedores conspiram com clientes no sentido que estes emitam excelentes avaliações de desempenho em troca de vantagens em transações futuras.

avaliações desonestas negativas (*bad-mouthing*): novamente os fornecedores e clientes conspiram de modo a mal afamar um determinado vendedor através de avaliações excessivamente baixas.

discriminação positiva e negativa: comportamento discriminatório de fornecedores em relação a clientes específicos. Para grupos específicos de clientes são fornecidos serviços de excelente qualidade e para o restante de média ou baixa qualidade. Esta discriminação pode ser cuidadosamente executada de forma a reputação do fornecedor não ser abalada –

se a quantidade de clientes discriminados for pequena ou se o fornecedor gozar de uma reputação acumulativa muito grande e começar a gradualmente alternar seu comportamento de modo a controlar a sua reputação.

transações falsas e inundação de avaliações desonestas: as estratégias acima podem ser viabilizadas e potencializadas através da criação de falsas transações que contenham valores de reputação escolhidos especificamente para tentar manipular, para cima ou para baixo, a reputação de alvos selecionados. A inundação também depende de transações falsas, mas neste tipo de ataque a manipulação ocorre por força bruta, e tem como objetivo específico aumentar a presença de avaliadores desonestos no sistema e aumentar o alcance das manipulações.

2.2.3.2 Classificação de Sistemas de Reputação

Vários esquemas de classificação [10, 26, 31] para sistemas e modelos de reputação vêm sendo propostos. A classificação proposta por Mui et al [10] é mais voltada para a definição de uma tipologia de reputação, que classifica reputações em individuais e de grupo, diretas e indiretas, baseadas em observação ou em encontros, propagadas ou previamente obtidas. O trabalho de Jøsang [26] será apresentado em maiores detalhes, em função de ser um dos mais atuais e abrangentes, e por ser organizado em torno do método (tarefa) de reputação e de aspectos técnicos de sistemas de reputação, que são assuntos recorrentes neste trabalho. As categorias de classificação são Semânticas de Confiança, Arquiteturas de Rede dos Sistemas de Reputação, e Mecanismos de Computação de Reputação.

Categoria Semânticas de Confiança

Os autores afirmam que as características semânticas das avaliações e das medidas de reputação e de confiança são importantes para que os participantes do sistema possam interpretá-las de forma adequada. Definem duas dimensões para a descrição da semântica:

- **Específico x Genérico:** a medida específica consiste em avaliar a confiança com base em um aspecto específico como por exemplo, a pontualidade da entrega do produto. A medida genérica é uma combinação de todos os aspectos específicos.
- **Subjetivo x Objetivo:** a medida subjetiva é obtida através de julgamentos subjetivos, enquanto a medida objetiva é obtida através de análises que utilizem critérios formais.

Os autores descrevem questionários de pesquisa como exemplos de instrumentos específicos e subjetivos, uma vez que cada questão avalia um aspecto específico, mas sobre a percepção subjetiva do respondente. Já testes de produtos são considerados como instrumentos específicos e objetivos, uma vez que várias unidades de medidas e marcas comparativas (*bench-marks*) podem ser usadas para avaliar aspectos como economia, rapidez, desempenho financeiro, etc.

Como exemplos de medidas genéricas e subjetivas classificam o sistema de reputação do eBay, em função das medidas de reputação não indicarem em que ponto o produto não atendeu às expectativas do comprador (entrega, avariado, etc.), e serem baseadas no julgamento do comprador. Como exemplo de genéricas e objetivas citam relatórios sumarizados de testes de produtos, que agregam medidas objetivas de cada característica do produto em uma pontuação única.

Categoria Arquiteturas de Rede dos Sistemas de Reputação

Nesta categoria os autores classificam os aspectos técnicos dos sistemas de reputação. As arquiteturas de rede se dividem em Centralizadas e Distribuídas, e são importantes por representarem como as informações de avaliação e de reputação fluem entre os participantes do sistema.

- **Arquiteturas centralizadas:** Consistem em sistemas de reputação centralizados, onde as avaliações de desempenho das entidades são fornecidas a uma autoridade central, com base na experiência direta dos membros da comunidade com as entidades avaliadas em transações realizadas. A autoridade central é responsável por produzir e manter atualizadas as medidas

de reputação de cada entidade avaliada com base nas avaliações recebidas, e por fornecê-las mediante solicitações feitas por outros participantes que desejem decidir se iniciarão transações com as entidades escolhidas. Os dois principais aspectos desta arquitetura são os protocolos de comunicação centralizados, que permitem informar avaliações e solicitar as medidas de reputação à autoridade central, e o mecanismo de computação de reputação, responsável por agregar as avaliações de desempenho em uma medida de reputação.

- **Arquiteturas distribuídas:** Em sistemas de reputação distribuídos não há possibilidade de manter um nó central para coleta de avaliações e computação de reputação. As informações de avaliação ficam distribuídas em vários locais. Um membro da comunidade, que deseja verificar a reputação de um terceiro, deve localizar estes locais distribuídos ou requisitar avaliações a outros membros que tiveram experiências pessoais com o terceiro, e a partir daí, o próprio solicitante deve efetuar a computação da reputação. Os dois principais aspectos são: os protocolos de comunicação distribuídos, usados para requisitar avaliações de terceiros aos membros da comunidade, e um método para que a própria parte interessada derive medidas de reputação a partir das avaliações recebidas.

Domínios de sistemas ponto a ponto (*P2P*) suportam estes tipos de arquitetura. Neste contexto, dada a natureza *servent* de seus membros (fornecem avaliações e consomem produtos), os sistemas de reputação têm dupla função: apontar os membros que oferecem boa qualidade de serviço, e apontar os membros que fornecem boas informações de avaliação de reputação.

Categoria Mecanismos de Computação de Reputação

Existem várias formas de agregar as informações de avaliação dos membros em valores de reputação. O trabalho em [26] descreve várias abordagens, entre elas as baseadas em modelos como as Redes Bayesianas e os Modelos Nebulosos, que a partir de conjuntos parciais de avaliações derivam modelos utilizados para estimar reputação, e as baseadas em memória como

as somas, médias e integrais, que sempre atuam sobre todo o conjunto de avaliações mantido pelos mecanismos de reputação.

- **Soma ou média das avaliações:** sistemas como eBay mantêm valores acumulados de reputação, incrementando-os à medida que são recebidas avaliações positivas e diminuindo-os quando são recebidas avaliações negativas. Seu principal benefício é a simplicidade da computação e a facilidade de entendimento, entretanto, não refletem as diferenças de gosto dos avaliadores e os contextos das avaliações, e não refletem claramente o comportamento negativo de membros que se aproveitam de suas reputações acumuladas para exercer comportamento desonesto ou mal intencionado.

Em outros sistemas de comércio eletrônico (como o *Amazon*) e sistemas de reputação, são utilizadas médias dos valores de avaliação, e médias ponderadas com base em fatores como recentidade das avaliações [4], credibilidade dos avaliadores [11, 32], reputação dos avaliadores [13], similaridade de perfis entre o avaliador e o usuário ativo [4], etc.

Também destacamos um método formal para cálculo proposto em [33] para proteger o sistema de ataques e manipulações, que consiste no cálculo de integrais para maximizar o custo da oscilação de comportamento de membros maliciosos.

2.2.3.3 Desafios e áreas de pesquisas

Sistemas de reputação vêm sendo implementados com sucesso em sistemas de comércio eletrônico, onde grande parte das soluções se baseia em arquiteturas centralizadas de informações. Novos domínios de aplicação surgem à medida que tecnologias como redes ponto-a-ponto totalmente descentralizadas e sistemas multiagentes vêm se desenvolvendo. A natureza distribuída e heterogênea destes ambientes requer o desenvolvimento de arquiteturas de reputação distribuídas.

Neste contexto, o projeto de mecanismos de reputação que levem em consideração a convivência com outras soluções de reputação afins, a capacidade de conferir maior confiabilidade aos testemunhos, e a capacidade de manter as informações de reputação sempre

atualizadas e acessíveis a todos os membros da comunidade, e em tempos aceitáveis, são importantes temas de estudo abertos pela profusão dos ambientes distribuídos e heterogêneos.

Outras áreas apontadas em [34] envolvem estudos sobre: usabilidade dos sistemas de reputação para tornar as medidas de reputação mais facilmente entendidas pelos usuários; efeitos da aplicação de mecanismos de reputação em ambientes onde haja atuação competitiva entre seus membros; danos causados por sistemas de reputação a imagens de membros através da disponibilização pública de informações de reputação que não correspondam ao seu comportamento real.

2.3 Filtragem Colaborativa Automática aplicada aos Sistemas de Reputação

Sistemas de Reputação e Sistemas de Recomendação baseados na técnica de FCA compartilham várias similaridades, e por esta razão são frequentemente confundidos na literatura e na sua referência em sistemas comerciais. Além de usarem como insumo avaliações coletadas dos usuários do sistema, possuem atuação similar à medida que ambos procuram ajudar os clientes a obterem serviços e produtos de qualidade [26].

Entretanto, existem diferenças que se forem observadas justificam as classificações distintas. Os sistemas de reputação utilizam uma abordagem mais pessimista, que consideram os membros da comunidade desonestos, oferecendo baixa qualidade de serviço para obtenção de lucros. Neste sentido, visam evitar que os clientes interajam com maus fornecedores.

Já a técnica de FCA originalmente proposta parte da premissa de que todo membro do sistema é honesto e procura oferecer o melhor produto ou serviço possível. Desta forma, a FCA visa recomendar as melhores opções de serviços possíveis. Uma das formas principais de fazer isto é através da identificação de similaridades entre usuários (formação de vizinhanças) a partir da semelhança entre seus gostos, e usar isto para recomendar itens que seus vizinhos tenham gostado.

Outra diferença percebida é que as recomendações feitas pela técnica de FCA podem ser inesperadas, enquanto nos Sistemas de Reputação o usuário geralmente escolhe seus parceiros de interação e somente depois disto obtém o valor estimado de reputação dele.

Os Sistemas de Reputação originalmente propostos partiam da premissa que os itens deveriam ser avaliados de forma consistente por qualquer usuário, e por isso trabalhavam com avaliações de reputação invariáveis aos gostos do avaliadores, por exemplo: um arquivo com vírus possui má reputação para todos os usuários, independente dos seus gostos, mas um arquivo de um filme pode ser julgado de forma distinta dependendo dos gostos dos avaliadores e do usuário solicitante.

À medida que as avaliações de reputação passaram a ser sujeitas ao gosto dos avaliadores, novos mecanismos de reputação [4, 13], baseados em filtragem colaborativa que fossem capazes de reconhecer variações nos gostos dos clientes e fornecer estimativas de reputação adequadas ao seus gostos começaram a ser propostos. Isto se torna mais claro quando verificamos que a similaridade de gostos é um dos fatores determinantes para o estabelecimento de boas avaliações de reputação. Se um fornecedor vende um CD que é muito apreciado porque se encaixou com o gosto do cliente, o fornecedor se torna um pouco mais confiável sobre o ponto de vista deste cliente. Estes trabalhos destacam os benefícios de personalização que a FCA proporciona aos mecanismos de reputação, principalmente com relação à geração de estimativas de reputações mais precisas e confiáveis.

Conforme descrito anteriormente, a FCA opera sobre estrutura de dados denominada Matriz de Avaliações (Tabela 2.2). Um exemplo de uma Matriz de Avaliações no domínio de reputação é apresentado na Tabela 2.4.

Avaliador / Item avaliado	Cliente A	Cliente B	Cliente C
Fornecedor 1	1	4	1
Fornecedor 2	0	2	0

Tabela 2.4 – Matriz de avaliações de fornecedores em um sistema de reputação.

Ao se aplicar FCA em sistemas de reputação, os itens avaliados passam a ser os próprios usuários, e a avaliação consiste na confiabilidade percebida pelo avaliador para cada usuário através do seu desempenho na transação. A idéia é que se a vizinhança do usuário ativo confia em um novo fornecedor, então o usuário ativo também confiará. A geração da estimativa de reputação personalizada segue os mesmos moldes da equação (2) apresentada na seção “Sistemas de Recomendação”, e repetida abaixo:

$$(2) p_{a,i} = \bar{r}_a + \frac{\sum_{u=1}^m (r_{u,i} - \bar{r}_u) * w_{a,u}}{\sum_{u=1}^m |w_{a,u}|}$$

A diferença é que nesta equação, $p_{a,i}$ corresponde ao valor estimado de reputação do fornecedor i para o usuário ativo, \bar{r}_a é a média das avaliações de confiabilidade do usuário ativo, $r_{u,i}$ é a avaliação de confiabilidade do vizinho sobre o fornecedor, \bar{r}_u é a média das avaliações do vizinho.

Uma vez apresentada a estrutura geral de funcionamento da filtragem colaborativa em sistemas de reputação, descrevemos, em maiores detalhes a seguir, duas variações de mecanismos de reputação baseados em filtragem colaborativa que testamos nos experimentos do capítulo 5.

2.3.1 Estimativas de reputação personalizadas e Filtragem de Agrupamentos

O mecanismo de reputação proposto por Dellarocas [4] é voltado para comunidades de comércio on-line, que se caracterizam pelos participantes a priori não se conhecerem e pela ausência de um histórico de transações passadas entre quaisquer pares de participantes. O mecanismo prevê a utilização de técnicas de filtragem colaborativa para geração de estimativas de reputações personalizadas para cada comprador ativo, de modo a compensar os diferentes gostos dos avaliadores, aumentar a confiabilidade das estimativas, e reduzir os efeitos de avaliações positivas desonestas (*ballot stuffing*) e da discriminação positiva de compradores. No caso desta

abordagem somente são geradas estimativas de reputação de vendedores, com os compradores desempenhando o papel de avaliador.

O mecanismo básico de Filtragem de Agrupamentos, baseado no domínio de valor, visa também proteger o sistema contra compradores desonestos que procurem se parecer com um determinado usuário para posteriormente oferecer avaliações desonestas sobre um fornecedor que desejam beneficiar. Para evitar estas situações, o mecanismo básico combina a técnica tradicional de FCA com o Algoritmo Divisor de Agrupamentos. Este algoritmo consiste em particionar o conjunto de vizinhos selecionados pela técnica de FCA em dois grupos, um contendo avaliadores honestos e outro com os avaliadores desonestos, e calcular a reputação sobre as avaliações do primeiro grupo. O particionamento deve ser feito através de uma função de distância que leve em conta a última avaliação ou da média das avaliações dos vizinhos sobre a entidade avaliada.

Na primeira etapa do algoritmo é calculada a soma das distâncias de cada avaliador para as demais testemunhas. O avaliador mais distante, aquele cuja avaliação menos se assemelha aos demais, é transferido para o grupo de avaliadores desonestos. Na segunda etapa são transferidos para o grupo de desonestos os avaliadores que menos se assemelham ao grupo original e que mais se assemelham com o de desonestos. Nesta etapa o cálculo consiste na diferença entre as somas das distâncias de cada avaliador para o seu grupo original e a soma das distâncias para o grupo de desonestos. O avaliador que tiver a maior diferença será transferido para o grupo de desonestos. Esta etapa é repetida enquanto houver mais de um membro no grupo original, e enquanto a diferença calculada for positiva, ou seja, enquanto o elemento avaliado for menos semelhante ao grupo original e mais semelhante ao grupo de desonestos.

Um mecanismo avançado, baseado no domínio de frequência, é proposto com o objetivo de proteger o sistema contra a inundação de avaliações positivas desonestas. Este tipo de ataque consiste na geração de uma quantidade muito grande de avaliações positivas desonestas em um curto espaço de tempo, fazendo com que o mecanismo básico seja ineficaz e não consiga remover todos os compradores desonestos do grupo de avaliadores selecionados para o cálculo da

reputação. Este algoritmo consiste na aplicação de dois níveis de particionamento, o primeiro nível consiste na execução do algoritmo divisor descrito anteriormente sobre a última avaliação feita por cada avaliador ao vendedor. O segundo nível consiste em reaplicar este algoritmo nos dois grupos inicialmente formados, só que desta vez utilizando como função de distância a quantidade média de avaliações de cada avaliador para o vendedor que ocorreram em um intervalo de tempo definido no momento de iniciação do algoritmo, o que resultará na formação de quatro grupos. A quantidade média de avaliações é recalculada nos quatro grupos, e aquele com menor média geral será usado como base para o cálculo de reputação.

2.3.2 Peso de Relevância (Significance Weighting)

Herlocker et al. [5] propõe um método para calcular a quantidade de confiança a ser atribuída às correlações entre o usuário ativo e seus vizinhos. Propõe ajustes na técnica de FCA com a aplicação de um peso sobre o coeficiente de similaridade conferido a cada testemunha, com base na quantidade de itens em comum avaliados anteriormente com o usuário ativo. Este ajuste deve ser feito entre a etapa de cálculo de similaridade da vizinhança e o cálculo da recomendação propriamente dito, e parte da premissa que quanto maior for a quantidade de itens em comum avaliados anteriormente maior é o grau de confiança conferido à correlação com o vizinho. O fator de correção é:

$$fc = \begin{cases} n/50 & \text{se } n < 50 \\ 1 & \text{se } n \geq 50 \end{cases}$$

onde n é a quantidade de avaliações em comum entre o avaliador e o comprador.

Em seus experimentos definiu que se a quantidade n de itens em comum for inferior a 50, o coeficiente de similaridade é reduzido ao ser multiplicado por $n/50$. Se a quantidade for superior a 50 o coeficiente é mantido em 1 (a escala do peso vai de 0 a 1) não sofrendo nenhum tipo de ajuste.

Utilizando como base a equação (2) apresentada na seção “Sistemas de Recomendação”, o cálculo da estimativa de reputação personalizada fica da seguinte forma:

$$(3) p_{a,i} = \bar{r}_a + \frac{\sum_{u=1}^m (r_{u,i} - \bar{r}_u) * w_{a,u} * f_{a,u}}{\sum_{u=1}^m |w_{a,u} * f_{a,u}|}$$

2.3.3 Conclusões

Apesar da personalização conferida pela FCA aos sistemas de reputação resultar em maior grau de confiabilidade das estimativas de reputação, os cenários de oscilação de comportamento das entidades avaliadas, citados na introdução deste trabalho, ainda podem criar dificuldades para o estabelecimento de estimativas confiáveis. As abordagens revistas na seção 2.3, apesar de possuírem mecanismos para tratar problemas como falsa vizinhança e manipulação de avaliações, não levam em conta situações de oscilação de comportamento. Por esta razão, propomos um método que não permita esta oscilação criar situações de falsa vizinhança que resultem na geração de estimativas de reputação inadequadas para os usuários solicitantes.

Capítulo 3 Método Proposto

Com o objetivo de minimizar os efeitos negativos da oscilação de comportamento dos fornecedores sobre a confiabilidade das recomendações, expostos na introdução deste trabalho, é proposto um método que ajuste o histórico de avaliações com base na similaridade dos quesitos usados pelos avaliadores, suportado pelo modelo geral apresentado na Figura 3.1 e descrito na seção 3.1.

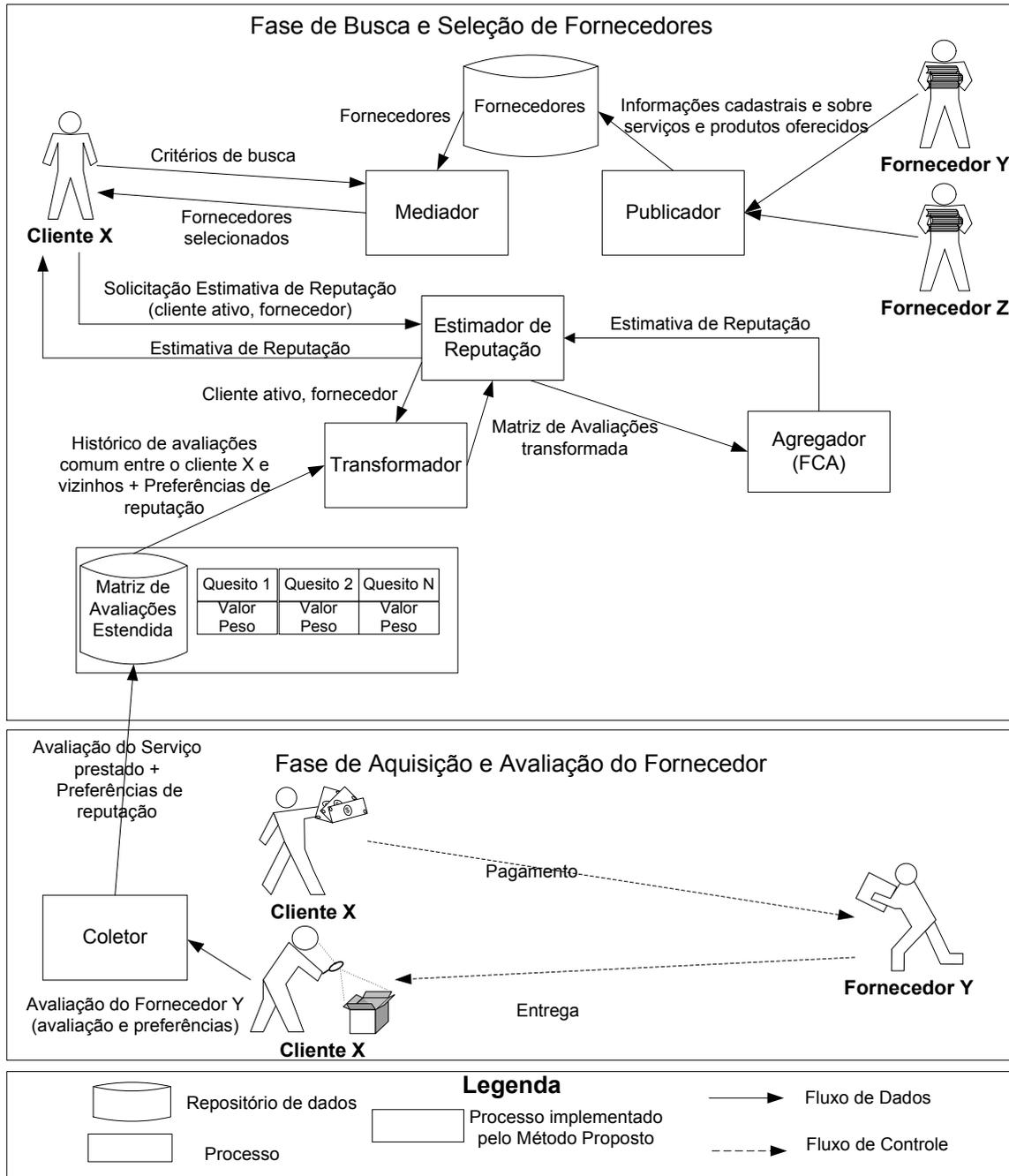


Figura 3.1 – Modelo geral do método proposto

3.1 Modelo de alto nível do método proposto:

As contribuições deste trabalho se refletem nas novas estruturas de dados criadas: Matriz de Avaliação Estendida e Preferências de Reputação, bem como nas novas funções desempenhadas através dos módulos Estimador de Reputação, Transformador e Coletor, descritas a seguir:

1. **Mediador e Publicador:** O cliente ativo utiliza quaisquer recursos de mediação disponíveis para localizar e negociar com fornecedores e solicita o valor estimado de reputação sobre o fornecedor escolhido. Na Figura 3.1 poderia ser utilizado um mediador muito utilizado no domínio de sistemas multiagentes, do tipo *Matchmaker* [35], e o protocolo de negociação do tipo Leilão (*Auction*), mas não há restrição de utilização de outros mecanismos. Para serem localizados, os fornecedores precisam informar por intermédio de um publicador informações cadastrais e sobre seus produtos e serviços.
2. **Estimador de Reputação:** Responsável por receber solicitações de estimativa de reputação dos clientes, e por coordenar o trabalho dos módulos Transformador e Agregador para geração das estimativas solicitadas.
3. **Transformador:** Recupera todo o histórico de avaliações em comum entre o cliente ativo e outros clientes (vizinhos). Também recupera as preferências de reputação utilizadas por cada avaliador em seu testemunho. Para minimizar os problemas de oscilação de comportamento, antes do Agregador calcular a similaridade de cada vizinho, o Transformador ajusta a Matriz de Avaliação de acordo com a similaridade das preferências de reputação dos avaliadores e o usuário ativo. Cada célula da matriz é recalculada conforme descrito na seção “Transformação da Matriz de Avaliação”, dando origem à matriz transformada.
4. **Agregador:** Sobre a matriz transformada aplica o algoritmo tradicional de FCA descrito no capítulo 2, que gera o valor preditivo de reputação do fornecedor escolhido. O algoritmo está configurado com o método de correlação usuário-a-usuário originalmente

proposto por Resnick em [1], com a medição de similaridade entre vizinhos através do Coeficiente de Pearson, a seleção de vizinhança por número de vizinhos, e geração de recomendações através de predição. A contribuição do método proposto é apenas no insumo do algoritmo, que passa a ser a matriz de avaliações transformada.

5. **Coletor:** Ao término da transação o cliente testemunha sobre o desempenho do fornecedor. Diferente das abordagens tradicionais, cada testemunho deve conter as avaliações atribuídas por quesito de avaliação e também as preferências do avaliador.

3.2 Transformação da Matriz de Avaliação e cálculo de reputação

Esta seção descreve como ocorrem as transformações sobre a matriz de avaliações implementadas pelo módulo Transformador apresentado na seção 3.1. Inicialmente define a estrutura de dados Matriz de Avaliações e posteriormente detalha a transformação propriamente dita através de equações e exemplos.

3.2.1 Matriz de Avaliação estendida

Para que a transformação, indicada no módulo Transformador do modelo de alto nível, possa ser realizada é necessário trabalhar com uma Matriz de Avaliação Estendida, que contemple além dos próprios valores de reputação, os gostos dos avaliadores e os valores de reputação atribuídos por quesito de avaliação (item do contrato). Os gostos dos avaliadores são representados como Preferências de Reputação, que são estruturas de dados inspiradas nos Aspectos Comportamentais ϕ e Estruturas Ontológicas definidas no sistema ReGreT [32]. A representação conceitual da Matriz de Avaliação Estendida bem como das Preferências de Reputação podem ser vistas no diagrama de classes da Figura 3.2.

Como pode ser visto na Figura 3.2, os clientes e fornecedores são entidades devidamente registradas no ambiente que realizam transações com objetivo de atendimento de necessidades financeiras e de consumo. Ao final de transação o cliente emite uma avaliação do fornecedor. Cada avaliação está associada a um contrato e à preferência de reputação vigente naquele momento. Os contratos são estabelecidos entre clientes e fornecedores para registrar as bases

acordadas (valores) para realização das transações. Cada contrato é constituído por cláusulas associadas aos itens acordados, como preço, prazo de entrega, quantidade e qualidade.

As preferências de reputação mudam ao longo tempo e são empregadas pelo cliente na avaliação dos fornecedores contratados. Elas são compostas por quesitos de avaliação. Cada item de contrato está associado a um quesito de avaliação que possui um peso e forma de avaliação. Os quesitos de avaliação indicam como os desvios dos valores finais em relação aos valores acordados nas bases do contrato influenciam negativamente ou positivamente em cada aspecto comportamental. Neste sentido têm função similar às Relações de Base (*Ground Relation*) definidas no sistema ReGreT [32], sendo a influência apurada através de uma função de avaliação dependente do domínio e descrita no atributo expressão da classe FuncaoAvaliacao.

O método proposto compartilha a premissa do ReGreT de que a reputação é um conceito complexo (Aspectos Comportamentais Complexos) aferido através da combinação de várias dimensões de avaliação mais simples (Aspectos Comportamentais Simples). Desta forma cada preferência de reputação é composta por vários quesitos de avaliação, que combinados com seus pesos determinam o valor de reputação geral, formando um relacionamento similar às Estruturas Ontológicas definidas no sistema ReGreT [32]. No exemplo da Tabela 3.1, o aumento do preço tem maior influência negativa (0,6) em relação à influência positiva (0,4) de se aumentar a qualidade sobre o Aspecto Comportamental Bom_Vendedor.

Preferências de Reputação (Aspecto Comportamental Complexo)	Quesito de Avaliação (Aspecto Comportamental Simples)	Item	Peso	Influencia
Bom_Vendedor	Oferece_Preço_Caro	Preço	0,6	Negativa
	Oferece_Boa_Qualidade	Qualidade	0,4	Positiva

Tabela 3.1– Exemplos de Preferências de Reputação

A Tabela 3.2 é um exemplo de Matriz de Avaliação Estendida. As preferências de reputação se encontram na linha peso, enquanto os valores de reputação atribuídos por quesito de avaliação encontram-se na linha avaliação. A avaliação relativa é o produto do peso do item e a avaliação de reputação do item, e a avaliação geral é a soma das avaliações relativas.

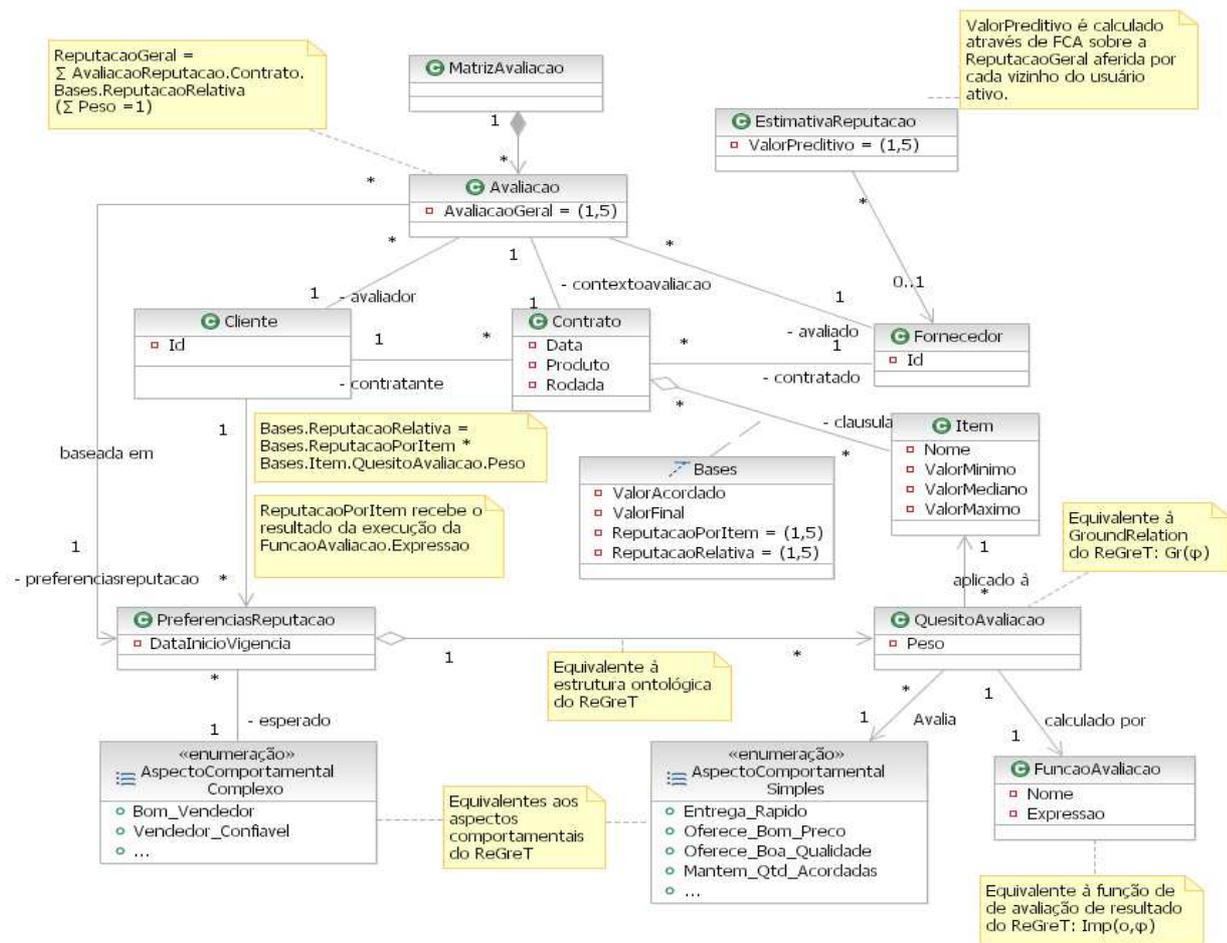


Figura 3.2 - Modelo conceitual completo (notação UML 2.0)

	Avaliador	Cliente 1				Cliente 2				Cliente 3			
Avaliado	Item	Qualidade	Preço	Data	Geral	Qualidade	Preço	Data	Geral	Qualidade	Preço	Data	Geral
Fornecedor X	Peso	0.3	0.2	0.5		0.2	0.3	0.5		0.5	0.3	0.2	
	Avaliação	5	4	3		4	5	3		3	5	4	
	Avaliação Relativa	1.5	0.8	1.5	3.8	0.8	1.5	1.5	3.8	1.5	1.5	0.8	3.8
Fornecedor Y	Peso	0.3	0.2	0.5		0.3	0.2	0.5		0.3	0.2	0.5	
	Avaliação	3	1	0		4	4	2		5	0	1	
	Avaliação Relativa	0.9	0.2	0	1.1	1.2	0.8	1	3	1.5	0	0.5	2
Fornecedor Z	Peso	0.8	0.1	0.1		0.2	0.3	0.5		0.5	0.3	0.2	
	Avaliação	5	4	3		5	5	4		5	4	3	
	Avaliação Relativa	4	0.4	0.3	4.7	1	1.5	2	4.5	2.5	1.2	0.6	4.3
Fornecedor W	Peso	0.3	0.2	0.5		0.2	0.3	0.5		0.3	0.2	0.5	
	Avaliação	4	3	4		5	5	5		3	2	1	
	Avaliação Relativa	1.2	0.6	2	3.8	1	1.5	2.5	5	0.4	0.9	0.5	1.8

Tabela 3.2 – Exemplo de Matriz de Avaliação Estendida

3.2.2 Forma de realização da transformação

As equações (4) (5) e (6) implementam os cálculos necessários para avaliação de reputação bem como as transformações da Matriz de Avaliação, introduzidas pelo método proposto, e indicadas no módulo Transformador do modelo geral.

$$(4) \mathbf{Ra}_{vizinho, fornecedor} = \mathbf{R}_{clienteAtivo, fornecedor} * fs_{clienteAtivo, vizinho, fornecedor}$$

$$(5) R_{clienteAtivo, fornecedor} = \frac{\sum_{item} (a_{fornecedor, quesito} * P_{clienteAtivo, quesito})}{\sum_{quesito} (P_{clienteAtivo, quesito})}$$

$$(6) fs_{clienteAtivo, vizinho, fornecedor} = \text{co-seno}(\text{preferencias}_{clienteAtivo, fornecedor}, \text{preferencias}_{vizinho, fornecedor})$$

Onde,

$\mathbf{Ra}_{vizinho, fornecedor}$ – novo mecanismo introduzido por este método, representa o valor de avaliação geral da matriz de avaliação transformado de acordo com o fator de similaridade entre o Cliente Ativo e o Vizinho. Os exemplos da Tabela 3.4 e Tabela 3.5 ilustram como ocorre a transformação da matriz de avaliação.

$\mathbf{R}_{clienteAtivo, fornecedor}$ - representa a avaliação geral do fornecedor sob a perspectiva do cliente ativo. Consiste na média ponderada das avaliações do cliente ativo sobre cada item do contrato e não em um valor único de reputação como ocorre em outros sistemas como Amazon [2], eBay [36], Internet Movie Database [37], e MovieLens [38]. A forma de cálculo independe do domínio de aplicação e da quantidade de itens de contratos negociados, e sempre resulta em valores na escala crescente de números reais entre 1 e 5.

$a_{fornecedor, quesito}$ - representa a avaliação do fornecedor de acordo com um determinado quesito de avaliação (item do contrato). A forma de cálculo depende do domínio de aplicação e do Aspecto Comportamental avaliado - exemplos são descritos nas equações (7) e (8) do capítulo 4 - sendo descrita pela Classe FuncaoAvaliacao apresentada anteriormente, mas deve produzir valores entre 1 e 5 para não comprometer o cálculo de reputação do fornecedor da equação (5).

$P_{clienteAtivo,quesito}$ - representa o peso e a importância dada pelo cliente ativo ao quesito. O peso é um número real entre 0 e 1 atribuído pelo cliente e que pode variar ao longo do tempo. A soma dos pesos dos itens deve ser sempre igual a 1. Exemplo:

Cliente	Peso da Qualidade	Peso do Preço
A	0,60	0,40
B	0,33	0,77

Tabela 3.3 – Exemplo de pesos atribuídos por um cliente a cada quesito de avaliação

$fS_{clienteAtivo,vizinho}$ – outro mecanismo introduzido por este método, representa o fator de similaridade entre o cliente ativo e um determinado vizinho. É determinado através de uma função de co-seno que calcula a distância entre os vetores de pesos dos quesitos da avaliação e assim identificar a semelhança das preferências de reputação empregadas pelo cliente ativo e seus vizinhos ao avaliar um fornecedor em comum.

3.2.3 Exemplo de transformação da matriz de avaliação

Considerando o Cliente 1 como sendo o cliente ativo, e aplicando as equações (4) (5) e (6) sobre a matriz de avaliação da Tabela 3.4 temos como resultado a Tabela 3.5 :

	Avaliador		
Avaliado	Cliente 1	Cliente 2	Cliente 3
Fornecedor X	3.8 (R)	3.8	3.8
Fornecedor Y	1.1 (R)	3	2
Fornecedor Z	4.7 (R)	4.5	4.3
Fornecedor W	3.8 (R)	5	1.8

Tabela 3.4 – Matriz de avaliação (resumida) do cliente ativo

	Avaliador		
Avaliado	Cliente 1	Cliente 2	Cliente 3
Fornecedor X	3.8	3.7 (Ra)	3.1 (Ra)
Fornecedor Y	1.1	1.1 (Ra)	1.1 (Ra)
Fornecedor Z	4.7	2.25 (Ra)	4.22 (Ra)
Fornecedor W	3.8	3.7 (Ra)	3.8 (Ra)

Tabela 3.5 – Matriz de avaliação do cliente ativo 1 transformada

Considerando as equações (4) (5) e (6), e a matriz estendida de exemplo (Tabela 3.2), a resolução da linha correspondente ao Fornecedor X na matriz transformada da Tabela 3.5 é demonstrada abaixo:

$$R_{cliente1,fornecedorX} = 3.8$$

$$fS_{cliente1,cliente2,fornecedorX} = \text{co-seno}(preferencias_{cliente1,fornecedorX}, preferencias_{cliente2,fornecedorX}) \Rightarrow$$

$$\text{co-seno}(0.3;0.2;0.5, 0.2;0.3;0.5) \Rightarrow \mathbf{0.973684}$$

$$fs_{cliente1,cliente3,fornecedorX} = \text{co-seno}(preferencias_{cliente1,fornecedorX}, preferencias_{cliente3,fornecedorX}) \Rightarrow$$

$$\text{co-seno}(0.3;0.2;0.5, 0.5;0.3;0.2) \Rightarrow \mathbf{0.815789}$$

$$Ra_{cliente2,fornecedorX} = R_{cliente1,fornecedorX} * fs_{cliente1,cliente2,fornecedorX} \Rightarrow 3.8 * 0.973684 \Rightarrow \sim = \mathbf{3.7}$$

$$Ra_{cliente3,fornecedorX} = R_{cliente1,fornecedorX} * fs_{cliente1,cliente3,fornecedorX} \Rightarrow 3.8 * 0.815789 \Rightarrow \sim = \mathbf{3.1}$$

A variação do comportamento do Fornecedor X entre as interações com os clientes 1, 2 e 3, fez com que as avaliações emitidas pelos clientes fossem iguais (3.8), apesar de terem Preferências de Reputação distintas. Este caso exemplifica o cenário do **falso bom vizinho** descrito na introdução deste trabalho, uma vez que o Cliente 3 acaba sendo tão bom vizinho do Cliente 1 quanto o Cliente 2. O ajuste feito na transformação corrigiu esta distorção, e as diferenças entre os clientes 2 e 3 em relação ao cliente 1 foram refletidas mais adequadamente na matriz resultante.

No caso do fornecedor Y, a mudança de comportamento que apresentou entre as interações com os clientes 1,2 e 3 tornou as avaliações emitidas pelos clientes diferentes (1.1, 3 e 2), apesar das preferências empregadas em cada avaliação terem sido iguais (0.3, 0.2, e 0.5). Este caso exemplifica o cenário do **falso mau vizinho** descrito na introdução, na medida que os clientes 2 e 3 se mostrem menos similares ao cliente 1 do que o são na realidade. Estas distorções também foram corrigidas na matriz resultante. O resultado dos ajustes realizados reflete no comparativo de apuração da vizinhança a seguir:

Vizinhança	Correlação c/ Cliente 1 (Correlação de Pearson)
Cliente 2	0.81
Cliente 3	0.65

Tabela 3.6 – Correlação entre os clientes na matriz de avaliação original do cliente ativo 1

Vizinhança	Correlação c/ Cliente 1 (Correlação de Pearson)
Cliente 2	0.66
Cliente 3	0.98

Tabela 3.7 – Correlação entre os clientes na matriz de avaliação ajustada do cliente ativo 1

Na matriz de avaliação original a opinião do Cliente 2 (0.81 de similaridade) em uma recomendação terá maior peso do que a opinião do Cliente 3 (0.65 de similaridade), em razão de ser mais similar ao cliente 1 do que o cliente 3. Mas ao se executar as transformações as correlações são alteradas

e o Cliente 3 (0.98) torna-se mais similar ao Cliente 1 do que o Cliente 2 (0.66). A inversão na ordem de importância dos vizinhos pode ser explicada com uma análise da Tabela 3.8: os valores de avaliação do Cliente 3 não são tão similares aos do Cliente 1, mas em contra-partida, as preferências de reputação dele se mantêm, ao longo das avaliações, mais próximas às do Cliente 1 do que as preferências do Cliente 2.

	Correlação c/ Cliente 1 ($f_{S_{clienteAtivo,vizinho,fornecedor}}$)	
	Cliente 2	Cliente 3
Fornecedor X	0.973684	0.815789
Fornecedor Y	1	1
Fornecedor Z	0.479234	0.898563
Fornecedor W	0.973684	1

Tabela 3.8 – Comparativo de similaridade de preferências com o Cliente 1

3.3 Propriedades do método proposto

Após descrever seus aspectos funcionais, destacamos a seguir algumas propriedades do método proposto que o tornam mais seguro, robusto e flexível. Acreditamos que as 3 primeiras propriedades especificamente criam as condições necessárias para melhor proteger o sistema contra manipulações por grupos de clientes e fornecedores desonestos, por causa do baixo nível de informação existente (propriedades distribuição e encapsulamento de informações) e das constantes transformações sofridas pela Matriz de Avaliação (propriedade tempo de vida e persistência das informações).

As estratégias de manipulação consistem na capacidade dos clientes emitirem avaliações similares a de outros clientes, aumentando a chance de se passar como vizinhos e de lucrar com as avaliações desonestas positivas (*ballot stuffing*) que realizarem posteriormente sobre outros fornecedores [4]. A forma de tornar as avaliações similares varia de acordo com a abordagem: em [4] os clientes monitoram as avaliações de clientes alvos, e posteriormente fazem suas próprias avaliações sobre os mesmos fornecedores com os mesmos valores dados pelo cliente alvo, ou inundam o sistema com avaliações desonestas positivas, já em [23], os valores de reputação são estimados com base nos valores centrais da distribuição de avaliações do sistema.

Distribuição das informações e funções: A responsabilidade de avaliação é distribuída pelos clientes. As Preferências de Reputação são mantidas pelos clientes. A coleta de testemunhos, a manutenção das Matrizes de Avaliações (estendida e transformadas), e a geração de recomendações são centralizadas no mecanismo de reputação.

Tempo de vida e persistência das informações: A Matriz de Avaliação Estendida é persistente e sofre atualizações à medida que as transações são finalizadas. As Matrizes de Avaliação transformadas são efêmeras e somente acessíveis no contexto do cliente ativo enquanto estiver usando o sistema. São sincronizadas com a Matriz de Avaliação persistente a cada nova solicitação de recomendação do cliente ativo.

Encapsulamento das informações: As avaliações de reputação por quesito e as Preferências de Reputação não são disponibilizadas para outros clientes, somente para o serviço de reputação. As recomendações geradas pelo Serviço de Reputação contêm apenas os valores estimados de reputação de cada fornecedor recomendado, e não contemplam informações sobre vizinhança nem peso das avaliações que originaram a recomendação.

Flexibilidade: O método foi elaborado de modo a poder conviver com outros modelos de reputação e de confiança baseados em fontes de informação diretas. Entretanto devido ao baixo nível de informação pública disponível, sua integração com outras abordagens baseadas em fontes de informação indiretas é dificultada (como redes sociais [32] e credibilidade de terceiros através da qualidade de seus testemunhos [11]). O método foi concebido de forma a ser empregado em domínios de aplicação como: Mercados Eletrônicos com interações conduzidas por agentes humanos [2, 36] ou computacionais [39], Sistemas ponto-a-ponto de compartilhamento de conteúdo parcialmente centralizados como Kazaa [40], desde que sejam mantidas as premissas de encapsulamento e de centralização de informações em serviços ou agentes de reputação.

3.4 Escopo de atuação

O escopo de atuação do método é basicamente sobre problemas relacionados à fase operacional de Agregação de *Feedback* descrita em [8]. Os problemas levantados nas fases de elicitação e distribuição de

feedback (mudança de identidade, pouca disposição dos participantes em reportar avaliações das partes, e garantia de testemunhos honestos) não são foco do presente trabalho, com exceção do problema relacionado à dificuldade em se reportar avaliações negativas, que é tratado através dos recursos de centralização e encapsulamento de informações discutidos anteriormente. Soluções para estes e outros problemas foram apresentados na seção “Sistemas de Reputação” do capítulo 2 deste trabalho.

Também não é escopo do método realizar o gerenciamento de contratos: verificar as evidências dos testemunhos, monitorar as interações entre as partes, garantir o cumprimento das cláusulas de contratos e identificar rompimentos ou desvios de cláusulas de contrato, nem tratar de forma especial contratos cancelados pelas partes. Soluções para isto já foram publicadas, tais como Agentes Notários [41], Notários e Monitores [42], e Monitores de Contrato e Serviços de Gerenciamento de Evidências [43].

O método proposto é dedicado somente à avaliação de fornecedores e utiliza todo histórico de avaliações feitas para cada fornecedor no cálculo de reputação. Algumas abordagens como *Sporas e Histos* [13] avaliam a reputação de clientes e fornecedores, utilizam as próprias reputações dos avaliadores no cálculo das reputações dos fornecedores, e selecionam apenas as avaliações mais recentes nos cálculos efetuados. Esta última característica se deve ao pressuposto de que as entidades avaliadas mudam de comportamento de forma linear ao longo do tempo, ou seja, que o comportamento mais recente não sofre oscilações significativas, e é a melhor referência para estimativa de comportamento futuro. No caso do método proposto, é levado em consideração que os fornecedores podem apresentar freqüentes oscilações de comportamento mesmo em períodos curtos de tempo. Por esta razão, se considera mais eficaz calcular o valor de reputação de um fornecedor com base na similaridade das preferências de avaliação entre seus avaliadores e o usuário ativo, do que descartar avaliações mais antigas, pois não há como garantir que somente as avaliações mais recentes reflitam adequadamente o comportamento atual do fornecedor.

Capítulo 4 Experimento

Para testar o método proposto foi desenvolvido um sistema que simula um mercado eletrônico de vendas de produtos, e registra os efeitos da oscilação de comportamento dos vendedores. O sistema foi desenvolvido em Java 1.5, com o algoritmo de FCA implementado através de um *framework* de filtragem colaborativa para Java denominado *Taste* [44]. A modelagem completa da simulação pode ser vista no anexo A deste trabalho.

O sistema é constituído por módulos que simulam usuários compradores (*buyers*) e vendedores (*sellers*), e por um serviço de reputação capaz de gerar recomendações baseadas nas várias técnicas de FCA testadas. Os compradores selecionam vendedores recomendados, seguem um protocolo simples de execução da transação e, ao término delas fornecem as avaliações dos vendedores ao serviço de reputação.

O serviço de reputação implementa o método proposto neste trabalho. No capítulo 5 são descritos experimentos comparativos onde o serviço de reputação passa a implementar mais três algoritmos de FCA testados. Todos os algoritmos compartilham configurações em comum, como o cálculo de similaridade entre vizinhos através do Coeficiente de Pearson, e a utilização do método de seleção de vizinhos por número de vizinhos (*best-n-neighbors*) [5, 14] configurado para 30 vizinhos.

4.1 Organização da Simulação

O experimento foi inspirado no trabalho conduzido em [45], incorporando a sua organização e forma de medir o desempenho dos métodos testados (através de recursos financeiros). A arquitetura do sistema é mais simples em razão do método proposto ser baseado exclusivamente na recuperação de testemunhos por mecanismos de Filtragem Colaborativa, enquanto o trabalho referenciado utiliza o sistema ReGreT, o qual se vale de outras dimensões e fontes de informações distintas para aferir os valores de reputação e de confiança.

As transações ocorrem ao longo de rodadas de treinamento e de teste. As rodadas de treinamento visam preparar a base de dados de reputação, e evitar que a esparsidade prejudique o desempenho dos algoritmos de FCA testados. Durante estas rodadas o módulo de reputação não fornece recomendações

(os vendedores são selecionados aleatoriamente), sendo apenas alimentado com as avaliações dos compradores.

As rodadas de teste são destinadas à monitoração do desempenho dos compradores e ao suporte às análises comparativas entre os métodos testados. Antes de iniciar as rodadas de teste, os compradores são reiniciados para se comportar como se tivessem chegado no sistema naquele momento. Passam a depender exclusivamente das estimativas de reputação geradas pelo sistema para selecionar vendedores e realizar transações, reduzindo a possibilidade de que qualquer efeito aleatório inviabilize a análise dos testes. A Matriz de Avaliações formada durante as rodadas de teste não é reinicializada para se assegurar a geração de estimativas de reputação através da Filtragem Colaborativa.

Os compradores selecionam vendedores com maior valor preditivo de reputação nas listas de recomendação geradas pelo módulo de reputação. Para permitir a análise da qualidade das recomendações geradas, os vendedores são configurados de modo a acatar as recomendações sem questioná-las, e a não empregar outros critérios para apoiar a seleção dos vendedores, tais como experiências diretas ou reputação do avaliador.

As listas de recomendação têm um tamanho máximo configurado de 5 posições. Cada comprador pode no máximo realizar uma transação por rodada. Cada vendedor pode participar de uma ou mais transações por rodada.

Durante as rodadas de teste o sistema não permite múltiplas interações entre um mesmo comprador e vendedor. Deste modo evita-se que sempre o mesmo vendedor (o primeiro da lista de recomendações) seja selecionado, garante-se a análise de toda a lista de recomendações (todos os vendedores das listas de recomendação de cada comprador são selecionados para transacionar), e permite-se a avaliação da evolução das recomendações ao longo das rodadas de teste.

Ao contrário dos experimentos inicialmente efetuados [46], cada simulação testa somente um tipo de algoritmo de FCA. Sendo assim, não há em uma mesma simulação, a divisão de compradores em grupos por algoritmo testado. Estes ajustes foram feitos em função de análises realizadas sobre os experimentos iniciais mostrarem que não era possível resguardar a validade dos testes estatísticos de

significância aplicados, em função da aleatorização [47] (estabelecimento das condições idênticas para cada grupo testado) não ser garantida: mesmo com a formação de grupos distintos de compradores utilizando algoritmos de FCA diferentes, ambos compartilhavam a mesma Matriz de Avaliação.

A medida que as recomendações de um tipo de algoritmo eram geradas e as interações entre compradores e vendedores ocorriam, novas avaliações eram inseridas na Matriz de Avaliação, que acabam sendo aproveitadas como insumos para o outro grupo de compradores testado. Os algoritmos de FCA se caracterizam por apresentar melhores resultados quanto maior for a quantidade informação disponível, sendo assim, a ordem de execução dos grupos de compradores foi determinante para o sucesso do algoritmo usado pelo grupo. Independente do tipo de algoritmo atribuído, o grupo que possuir os compradores que sejam executados por último tenderá a apresentar um desempenho melhor em relação aos demais.

Mesmo considerando que o cenário acima exposto possa ser mais próximo do mundo real (ambientes heterogêneos caracterizados pela convivência de diferentes algoritmos de FCA), o enfoque neste momento é na realização de análises comparativas imparciais, através de experimentos controlados e análises estatísticas, e por esta razão os algoritmos são testados separadamente.

4.2 Fases da Simulação

4.2.1 Fase de Negociação

Ao iniciar uma transação as partes acordam sobre o preço e a qualidade do produto. O preço está na faixa de valores entre 10,00 e 50,00, enquanto a qualidade assume valores naturais entre 1 (mais baixa) e 5 (mais alta). O acordo inicial é estabelecido com base nos valores medianos das faixas (30,00 para preço e 3 para qualidade).

4.2.2 Fase de Execução

Durante a transação o vendedor pode mudar os valores acordados com o comprador de acordo com seu tipo de comportamento, influenciando no resultado da transação. Existem três tipos de comportamento:

Mau (*Bad*) – Para simular as condições de oscilação de comportamento, durante as rodadas de treinamento, em 60% das transações aumentam o preço inicialmente acordado em $\frac{1}{4}$ e diminuem a qualidade em $\frac{1}{4}$. Em 40% das transações apresentam um comportamento equivalente ao Bom. Esta configuração faz com que estes vendedores sejam bem avaliados em algumas situações, e mal avaliados em outras. O percentual originalmente definido no experimento com o ReGreT [45] foi de 75%, entretanto neste caso faz mais sentido defini-lo em 60% para tornar as mudanças de comportamento mais frequentes. Durante as rodadas de teste, em 100% das transações aumentam o preço e diminuem a qualidade em $\frac{1}{4}$.

Estão configurados desta forma para reproduzir as situações de falsa vizinhança, mostradas na introdução deste trabalho, em que outros compradores com preferências semelhantes não sejam considerados como vizinhos, e vice-versa. Durante as rodadas de treinamento mudam constantemente de comportamento recebendo avaliações ruins e boas, e nas rodadas de teste se misturam aos agentes com bom comportamento no topo das recomendações geradas pelo módulo de reputação, aumentando a chance de decepção dos compradores que recebem as recomendações.

Bom (*Good*) - Ao longo de todas as rodadas aumentam a qualidade do produto em $\frac{1}{4}$, e diminuem o preço do produto em $\frac{1}{4}$. Estão configurados desta forma para beneficiar os compradores que efetivamente recebam as melhores recomendações, e deste modo representar um contra-ponto em relação à decepção que um comprador sofre ao receber uma má recomendação.

Neutro (*Neutral*) – Em todas as rodadas não efetuam mudanças sobre as cláusulas acordadas no contrato.

A quantidade de vendedores por tipo de comportamento varia de acordo com os cenários de teste, conforme apresentado mais adiante.

4.2.3 Fase de Encerramento

Ao final da transação os compradores executam três tarefas principais: avaliam o vendedor, informam o valor calculado ao módulo de reputação, e atualizam seus recursos financeiros.

4.2.3.1 Cálculo da Reputação do vendedor

Conforme descrito em detalhes no capítulo 3, a reputação do vendedor é calculada em uma escala crescente de números reais que vai de 1 a 5, sendo calculada como a média ponderada das avaliações dos itens do produto, conforme indicado na equação (5) ajustada para o domínio de simulação:

$$(5) R_{c,v} = \frac{(a_{v,qual} * p_{c,qual}) + (a_{v,preco} * p_{c,preco})}{(p_{c,qual} * p_{c,preco})}$$

onde, c representa o comprador, v o vendedor, a a avaliação do vendedor de acordo com os itens do produto, e p o peso dado pelo comprador ao item. A ponderação é feita sobre as preferências de reputação do comprador, que, no caso específico da simulação, representam o grau de importância conferida aos itens preço e qualidade. O peso é um número real entre 0 e 1 escolhido aleatoriamente para cada comprador no início da simulação, e que permanece fixo durante toda a simulação.

O método proposto prevê que a forma de avaliar os itens de contrato seja específica em cada domínio desde que o resultado calculado esteja na faixa de valores de 1 a 5. A avaliação por item é calculada de forma linear na simulação (poderia-se utilizar também funções não-lineares):

Item Qualidade: (7) $a_{v,qual} = vf_{v,qual}$

onde $vf_{v,qual}$ representa o valor final do item qualidade obtido através da Tabela 4.1.

Qualidade final do produto vendido ($vf_{v,qual}$)	Avaliação_Qualidade ($a_{v,qual}$)
5	5
4	4
3	3
2	2
1	1

Tabela 4.1 – Forma de obtenção de avaliação para o item qualidade

Item Preço: (8) $a_{v,preco} = 6 - (vf_{v,preco} / 10)$

onde $vf_{v,preco}$ representa o valor final do item preço conforme exemplificado na Tabela 4.2.

Preço final do produto vendido ($vf_{v,preco}$)	Avaliação_Preço ($a_{v,preco}$)
50,00	1
42	1,8
31	2,9
....
15	3,5
10,00	5

Tabela 4.2 – Exemplo de obtenção da avaliação para o item preço

4.2.3.2 Atualização dos recursos financeiros

Cada comprador inicia a simulação com recursos financeiros de 5000,00 que são atualizados ao final de cada transação conforme indicado nas equações (9) e (10):

$$(9) c_t = c_{t-1} - vf_{v,preco} + pr$$

onde c representa os recursos financeiros de que o comprador dispõe (c_t , os recursos da rodada atual e c_{t-1} , os recursos da rodada anterior), e pr representa o valor hipotético de revenda do produto comprado.

Foi definido que nas rodadas de teste os compradores efetuam a revenda dos produtos adquiridos, com o preço de revenda sendo determinado em função da qualidade do produto adquirido:

$$(10) pr = vf_{v,qual} * 10$$

sendo 10 o fator de conversão de preço.

A revenda resulta em lucro sempre que o comprador efetuar uma transação com um vendedor de bom comportamento, e resulta em prejuízo sempre que efetuar uma transação com um vendedor de mau comportamento. Com base nesta premissa é feita a comparação de desempenho entre os métodos testados no capítulo 5.

Os compradores selecionam os vendedores com base nas recomendações emitidas de acordo com o método definido para a simulação. Ao final de cada rodada é registrado o valor médio de recursos financeiros entre os compradores, e após o encerramento da simulação as distribuições de médias podem ser analisadas de forma a verificar se o método proposto é melhor do que os demais métodos testados, e se esta diferença é estatisticamente significativa.

4.3 Teste de geração de recomendações

Foi desenvolvido o teste de geração de recomendações com o objetivo de testar a hipótese 1, definida na seção 1.3, avaliando-se o desempenho do método proposto através da análise dos efeitos das recomendações geradas ao longo das rodadas de teste sobre os compradores.

O objetivo deste teste é analisar as recomendações geradas pelo método proposto, denominado FCA1. Para isto é verificado se os vendedores com comportamento bom ou neutro são recomendados antes dos vendedores com mau comportamento, e qual o impacto das recomendações sobre o lucro ou prejuízo apurado pela revenda dos produtos fornecidos pelos vendedores recomendados. Também é verificado como o método se comporta à medida que vendedores com perfil de mau comportamento se tornem predominantes no sistema. Vários cenários de teste foram previstos na simulação, representando diferentes proporções de vendedores por tipo de comportamento conforme a Tabela 4.3.

Quadro resumo simulações						
	Cenário I	Cenário II	Cenário III	Cenário IV	Cenário V	Cenário VI
Bad	40%	50%	55%	60%	65%	75%
Good	50%	40%	35%	30%	25%	25%
Neutral	10%	10%	10%	10%	10%	0%

Tabela 4.3 – Descrição dos cenários de simulação

Este teste usou a seguinte configuração:

- 90 rodadas, sendo 60 de treinamento e 30 de teste (aproximadamente 15% de esparsidade da matriz de avaliação);
- 13 transações por rodada;
- 13 compradores ;
- 64 vendedores;
- total de 1170 transações executadas.

4.3.1 Resultados

Os gráficos das figuras 4.1 a 4.6 demonstram o desempenho financeiro dos compradores por cenário de teste, descrevendo a média dos recursos financeiros dos compradores ao longo das rodadas de teste. Os recursos de cada comprador são calculados através das equações (9) e (10) descritas na seção “Atualização

dos Recursos Financeiros”. O crescimento da média ocorre por causa do lucro apurado resultante das transações com vendedores do tipo Good recomendados pelo serviço de reputação. Em contrapartida, o declínio ocorre devido ao prejuízo apurado com transações realizadas com vendedores do tipo Bad.

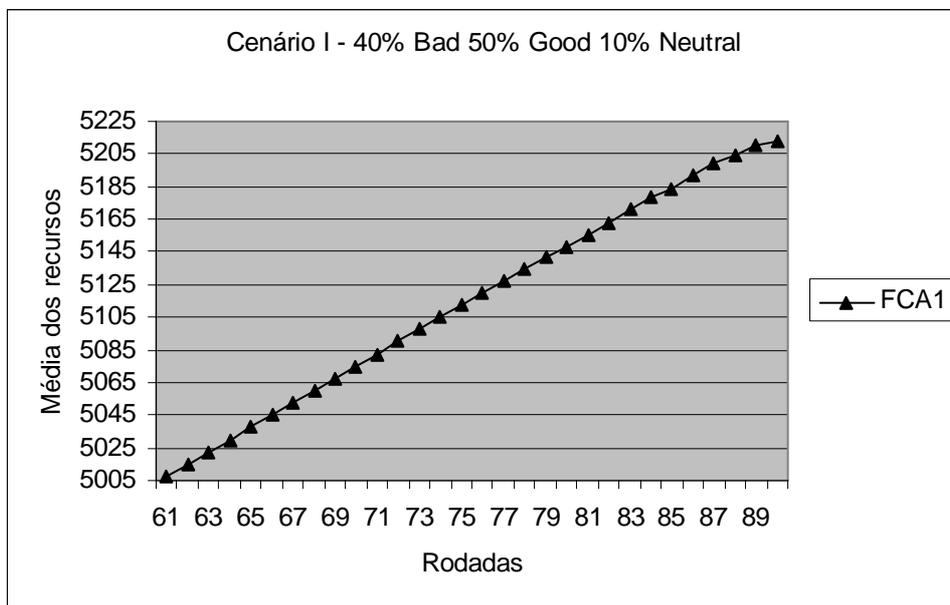


Figura 4.1 - Desempenho monetário dos compradores no cenário I

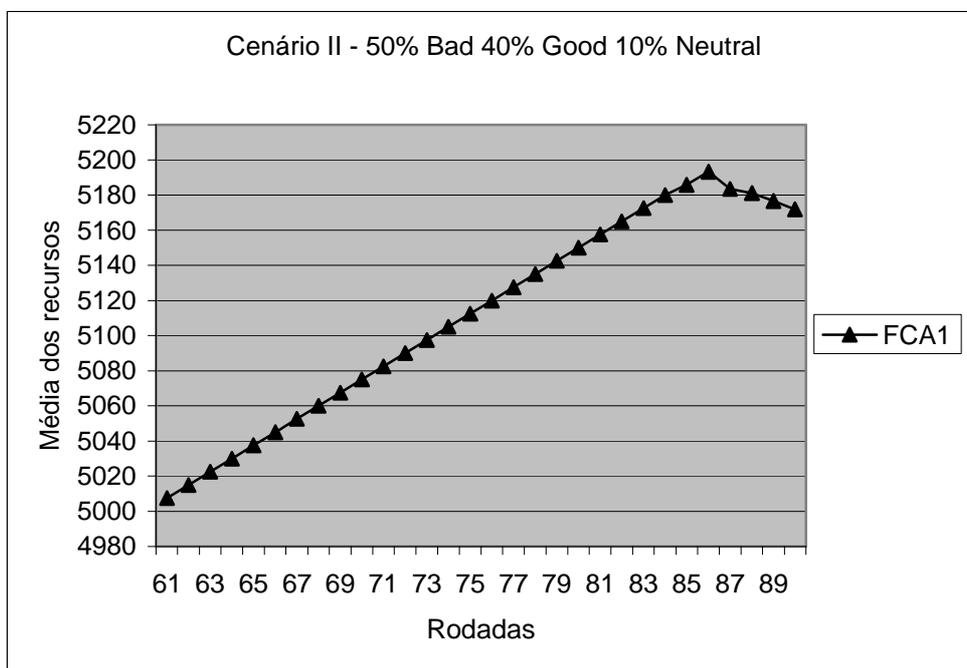


Figura 4.2 - Desempenho financeiro dos compradores no cenário II

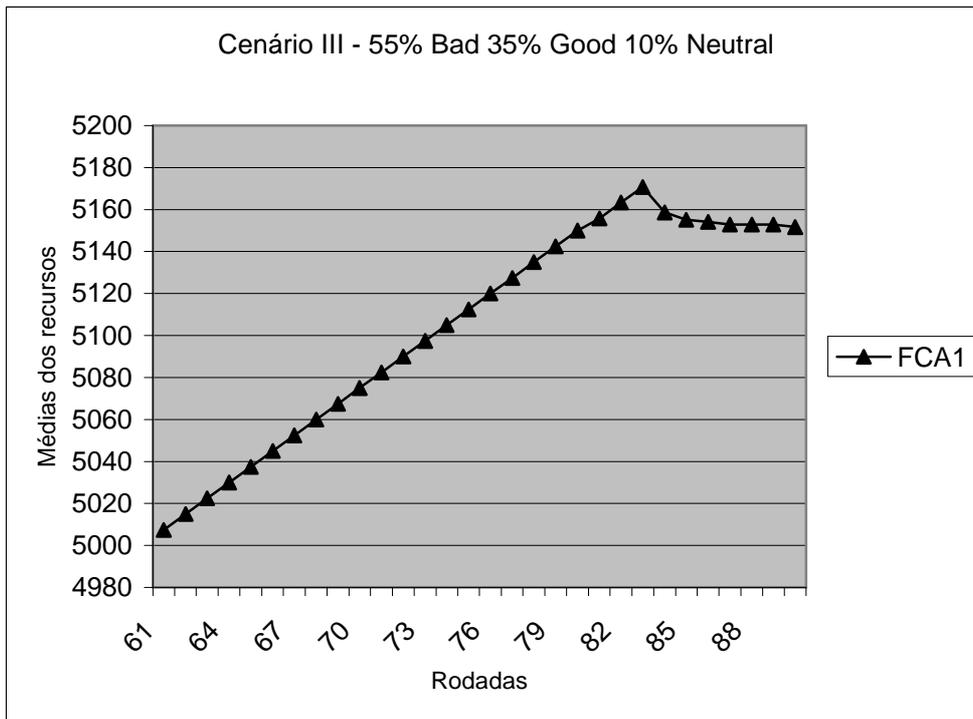


Figura 4.3 - Desempenho financeiro dos compradores no cenário III

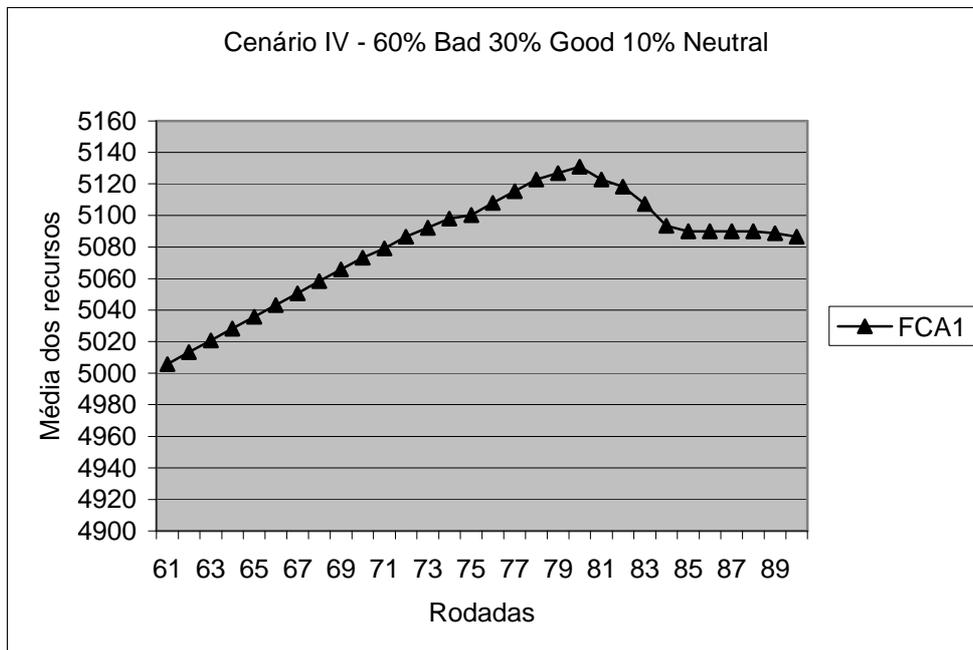


Figura 4.4 - Desempenho financeiro dos compradores no cenário IV

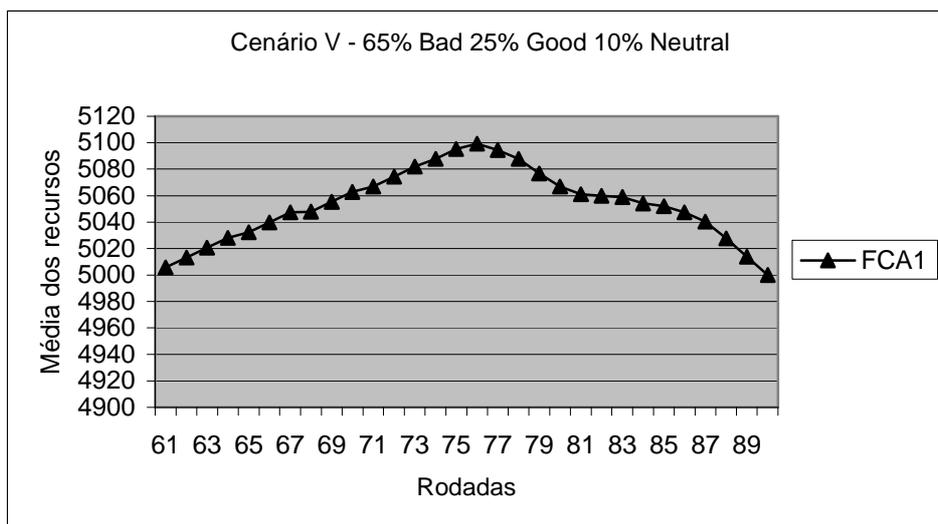


Figura 4.5 - Desempenho financeiro dos compradores no cenário V

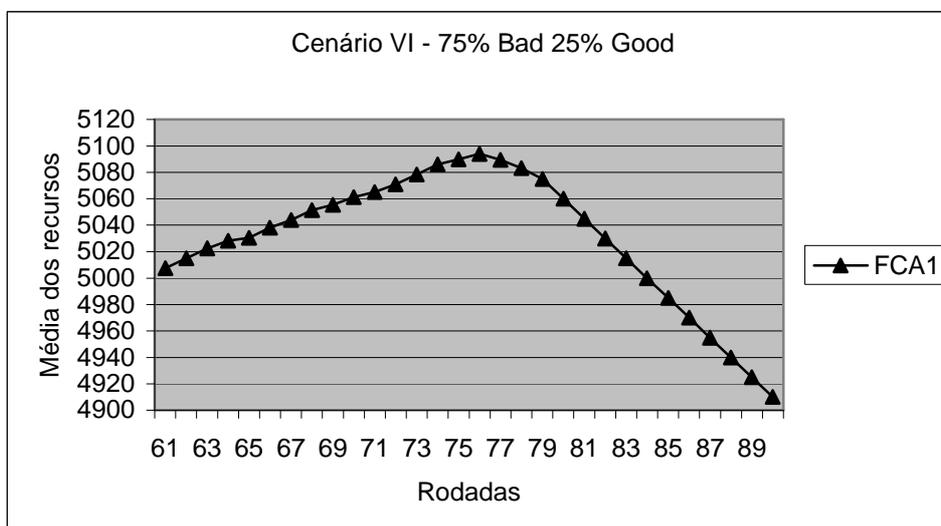


Figura 4.6 - Desempenho financeiro dos compradores no cenário VI

O padrão das curvas - crescimento contínuo nas rodadas iniciais seguido por um declínio menos acentuado no restante da simulação - mostra que de um universo de todos os tipos de vendedores, o método proposto (FCA1) gera inicialmente recomendações dos tipos Good, e posteriormente Neutral alternadas com Bad. Os gráficos representados nas figuras 4.7 a 4.12 confirmam esta afirmação. Este padrão somente não é observado no cenário I, que não apresenta declínio significativo em função da maior proporção de vendedores tipo Good do que Bad, e no cenário VI, que apresenta uma tendência de declínio mais acentuada do que os demais em função de não haver vendedores do tipo Neutral.

É possível verificar que a curva de declínio tende a se alongar à medida que aumenta a proporção de vendedores tipo Bad, iniciando na rodada 87 quando há 50% de tipo Bad (Figura 4.2), e na rodada 77 quando há 75% de tipo Bad (Figura 4.6).

A distribuição de recomendações também foi objeto de análise, conforme pode ser visto nos gráficos das figuras 4.7 a 4.12. Cada gráfico corresponde a um cenário e demonstra a distribuição das recomendações de vendedores de tipo Good, Bad e Neutral ao longo das rodadas de simulação.

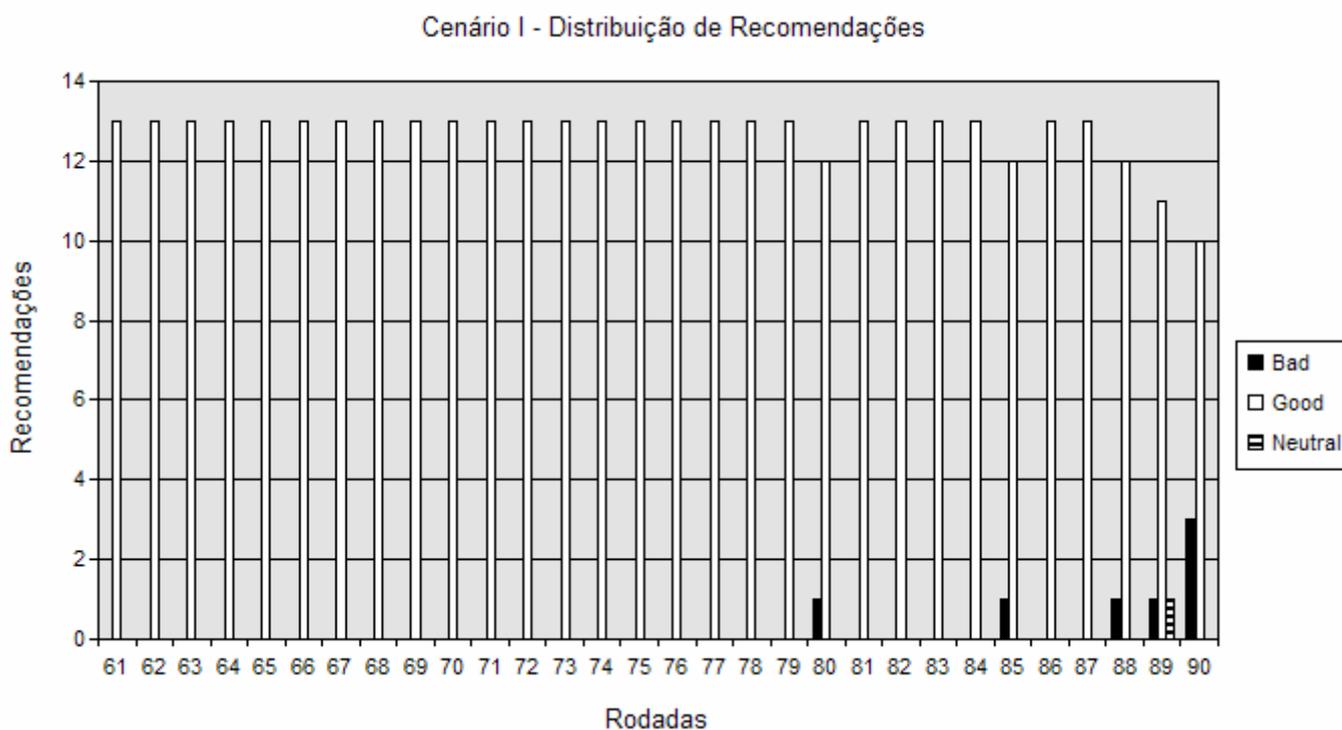


Figura 4.7 - Quantidade de recomendações de vendedores Bad, Good e Neutral por rodada do cenário I

Cenário II - Distribuição de Recomendações

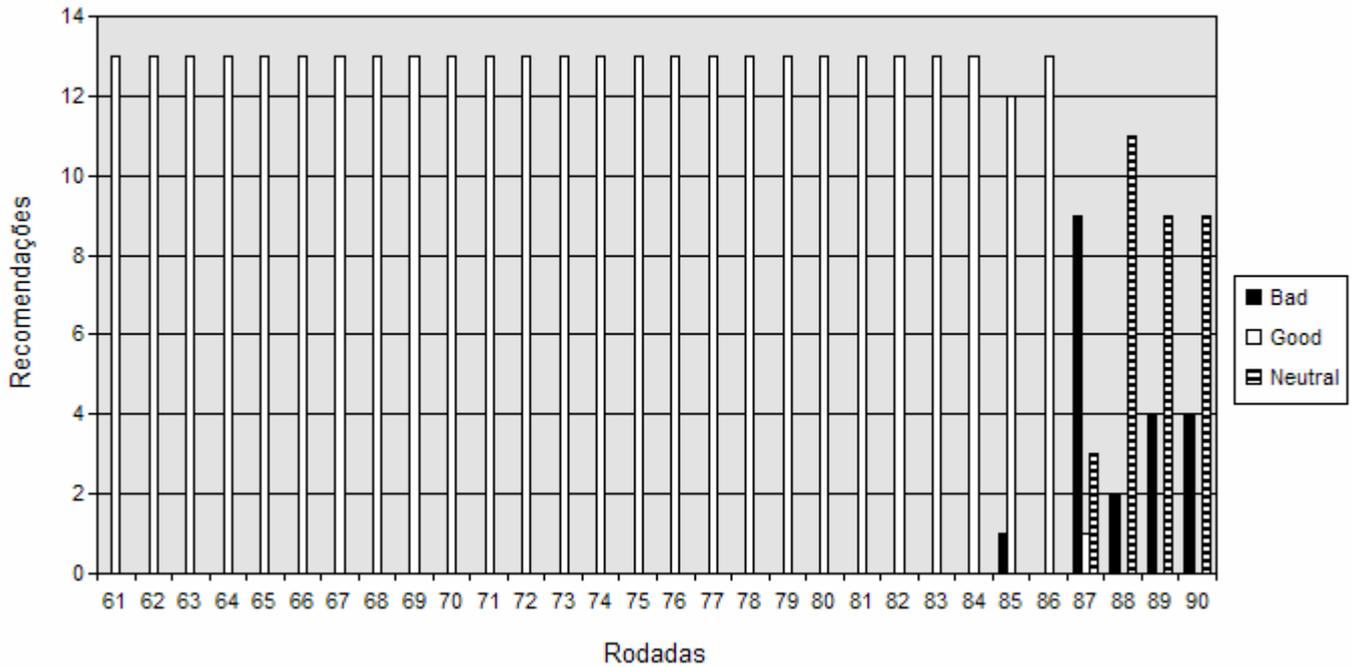


Figura 4.8 - Quantidade de recomendações de vendedores Bad, Good e Neutral por rodada do cenário II

Cenário III - Distribuição de Recomendações

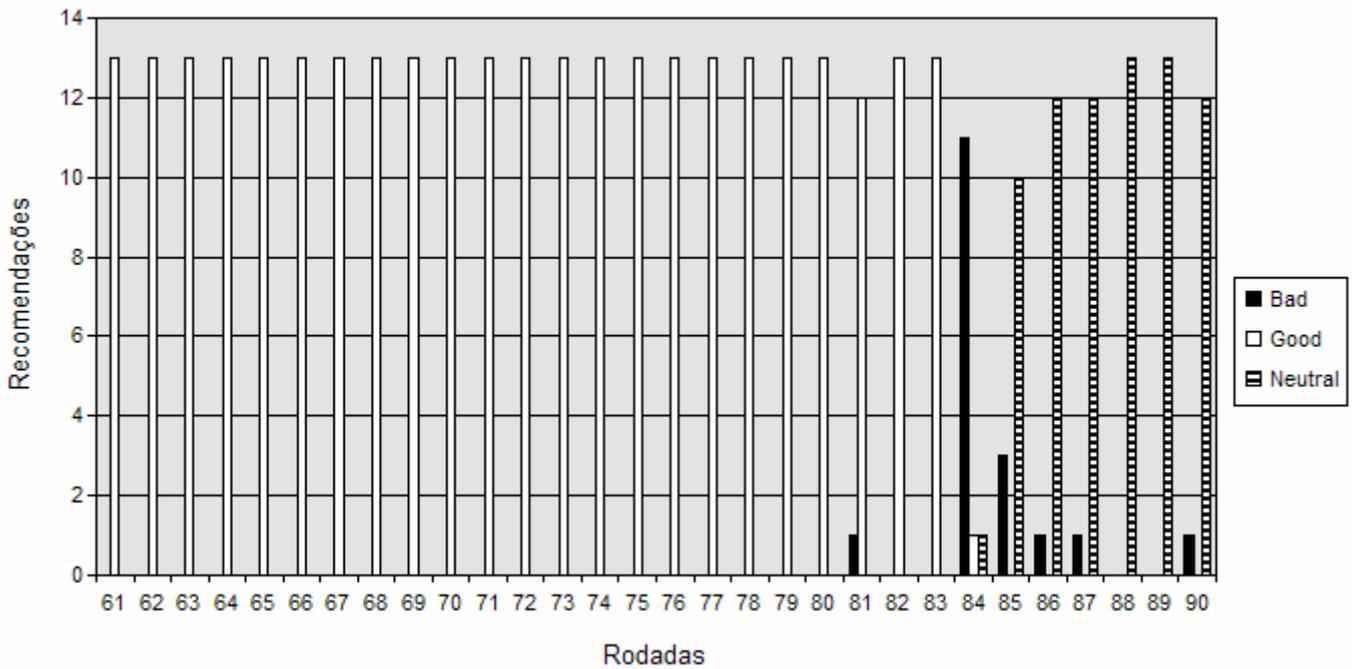


Figura 4.9 - Quantidade de recomendações de vendedores Bad, Good e Neutral por rodada do cenário III

Cenário IV - Distribuição de Recomendações

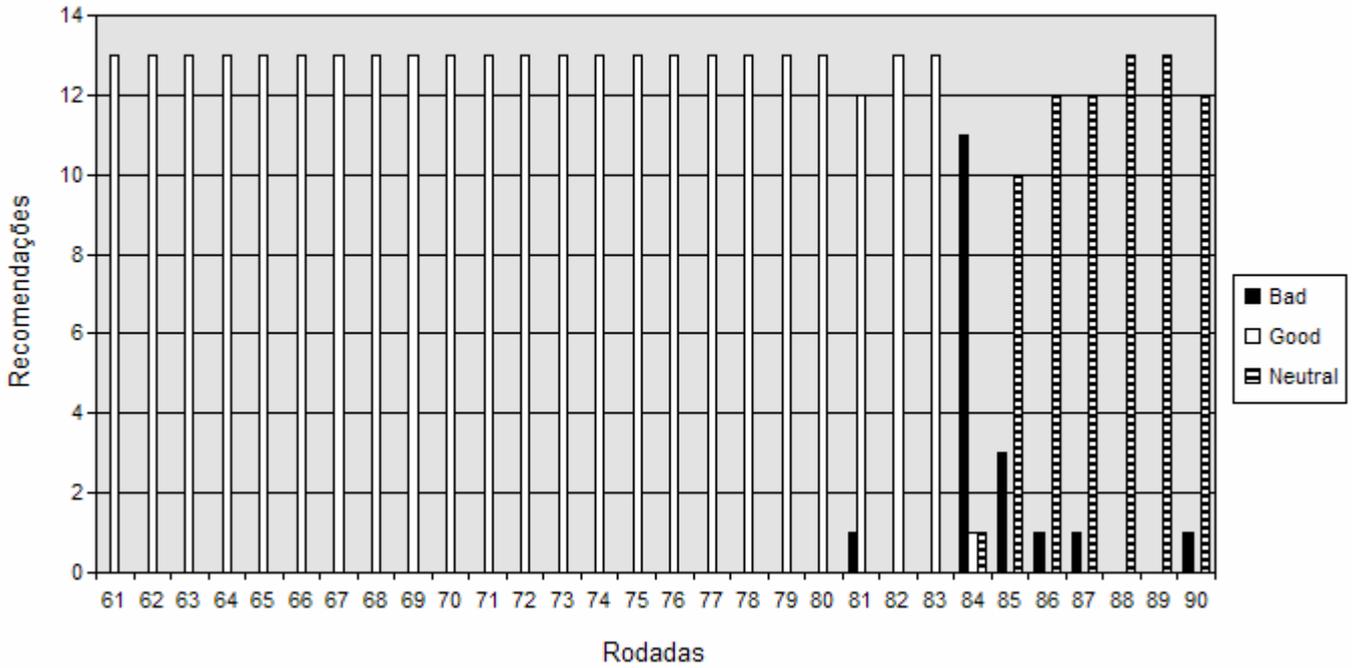


Figura 4.10 - Quantidade de recomendações de vendedores Bad, Good e Neutral por rodada do cenário IV

Cenário V - Distribuição de Recomendações

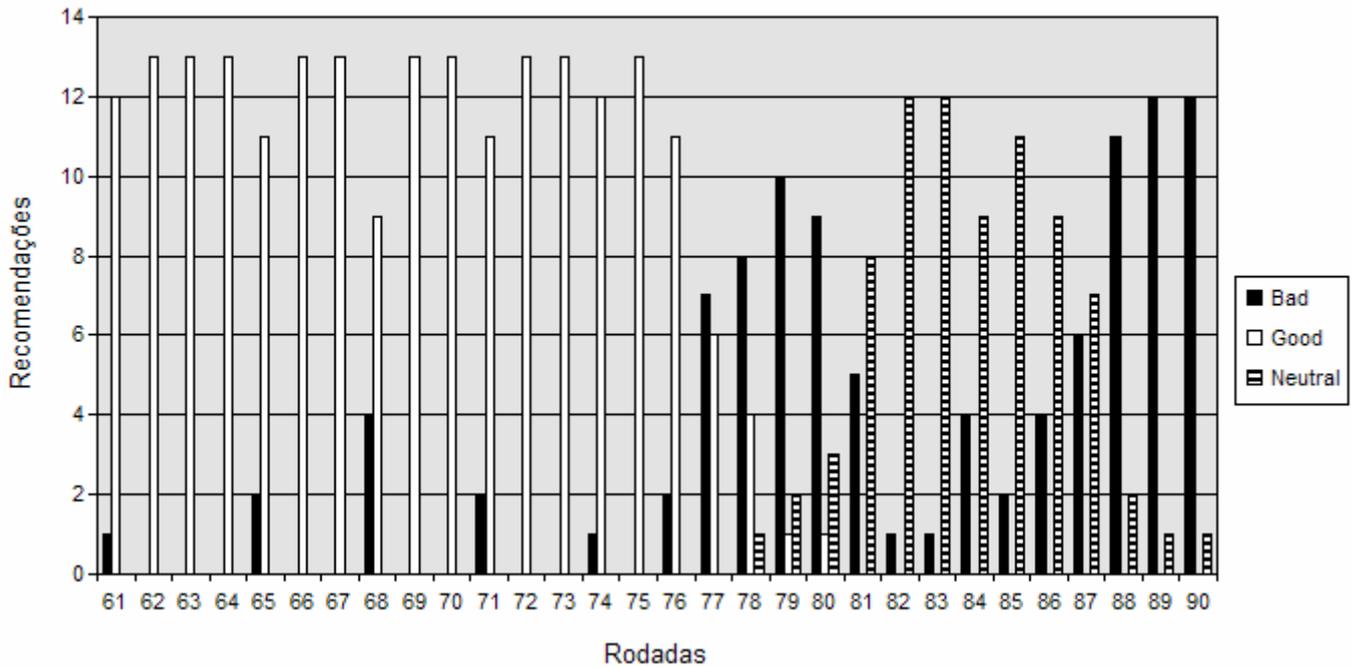


Figura 4.11 - Quantidade de recomendações de vendedores Bad, Good e Neutral por rodada do cenário V

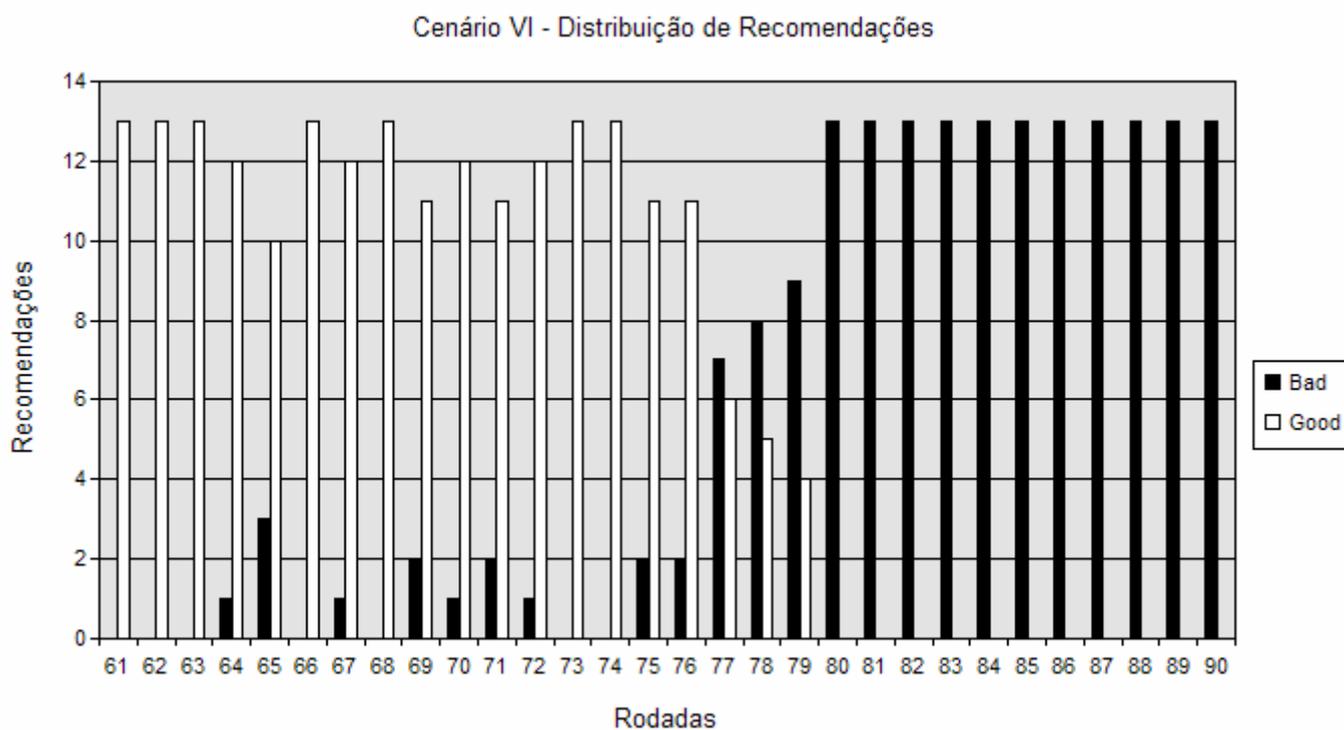


Figura 4.12 - Quantidade de recomendações de vendedores Bad, Good e Neutral por rodada do cenário VI

A análise dos gráficos das figuras 4.7 a 4.12 mostra que somente a partir do cenário V (65% Bads) há um número considerável de vendedores Bad que se infiltram entre as recomendações iniciais. Neste cenário já é possível observar vendedores Bad recomendados a partir da rodada 61, padrão que se acentuou nos cenários seguintes, mas não chegou a comprometer a média dos recursos financeiros, porque no cenário VI, de um universo de 208 recomendações geradas até a rodada 76, somente 15 eram de vendedores Bad (cerca de 7% do total).

Também fica evidenciado que as recomendações de vendedores Neutrals se intercalam às de vendedores Bad, o que explica o declive menos acentuado dos recursos financeiros observado nos gráficos da Figura 4.3 à Figura 4.5.

Por último, observou-se ao longo dos testes de recomendações que a média final dos recursos financeiros diminuiu com o aumento da proporção de vendedores tipo Bad, conforme mostrado no gráfico abaixo.

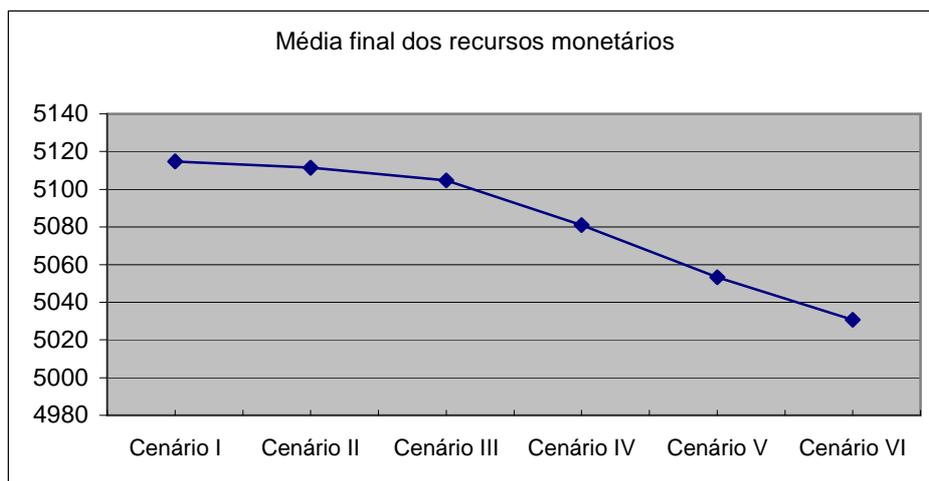


Figura 4.13 – Queda do desempenho financeiro ao longo dos cenários de teste

4.3.2 Conclusões

O método proposto neste trabalho se comportou adequadamente, mesmo em cenários de oscilação de comportamento por parte dos vendedores. Em todos os cenários priorizou recomendações de vendedores tipo Good, possibilitou a apuração de lucro aos compradores enquanto havia vendedores tipo Good disponíveis para transacionar, e pouco sofreu com a infiltração de vendedores Bad entre as recomendações iniciais.

Capítulo 5 Trabalhos correlatos

Neste capítulo estabelecemos uma análise comparativa entre o método proposto e trabalhos correlatos apresentados no capítulo Conceitos Fundamentais, baseada nos resultados obtidos nos testes comparativos descritos neste capítulo. Também são apresentadas outras abordagens que não foram testadas nos experimentos, assim como uma discussão sobre suas principais similaridades e diferenças para o método proposto.

5.1 Estimativas de reputação personalizadas e Filtragem de Agrupamentos

Conforme explanado na seção 2.3.1, o mecanismo de reputação proposto por Dellarocas [4] prevê a utilização de técnicas de FCA combinadas com mecanismos básicos (domínio de valor) e avançados (domínio de frequência), para geração de estimativas de reputações personalizadas, e para melhor proteger o sistema contra manipulações de entidades mal intencionadas.

O método proposto neste trabalho também utiliza a técnica de FCA para cálculo de reputações, só que sobre uma matriz de avaliação estendida, com as avaliações decompostas em diferentes quesitos de avaliação (preferências de reputação) e transformadas de acordo com a similaridade das preferências das testemunhas e do usuário ativo.

Conforme descrito na seção “Propriedades do Método Proposto” do capítulo 3, acreditamos que o encapsulamento e a transformação destas informações, promovidos pelo método proposto, sejam fatores diferenciais de proteção contra manipulações de avaliadores desonestos. Considerando que nem todas as informações utilizadas para o cálculo de reputação sejam de domínio público, a sua reprodução por avaliadores desonestos acaba sendo dificultada, tanto no domínio de frequência quanto no domínio de valores. Por estas razões o método proposto neste trabalho não se baseou nos mecanismos básico e avançado propostos por Dellarocas.

5.2 Peso de Relevância (Significance Weighting)

Herlocker et al. [5] propõe ajustes na técnica de FCA com a aplicação de um peso sobre o coeficiente de similaridade conferido a cada testemunha, com base na quantidade de itens em comum avaliados anteriormente com o usuário ativo.

O método proposto neste trabalho também funciona com a aplicação de pesos, só que aplicados diretamente sobre os valores de avaliação da matriz de avaliação, e não sobre coeficientes de similaridade dos vizinhos, que no método proposto, são calculados da mesma forma que na técnica tradicional de FCA. Os pesos consistem no cálculo do co-seno das preferências de reputação empregadas pelo usuário ativo e seus vizinhos nas avaliações de usuários em comum.

Ambos trabalhos minimizam os efeitos das coincidências sobre avaliações em comum, mas acreditamos que em domínios onde as recomendações são feitas sobre usuários e não produtos - como em sistemas de reputação - o problema de confiabilidade de testemunhas não é só determinado pela baixa quantidade de avaliações em comum, mas também pela oscilação de comportamento das entidades avaliadas. Por exemplo: é possível identificar um bom vizinho (mesmos gostos) mesmo com poucas avaliações e identificar um distante (falso bom vizinho) mesmo com muitas avaliações. Quanto maior a quantidade de avaliações maior a probabilidade de se deparar com usuários com comportamento oscilante, e maior a chance de se selecionar falsos bons vizinhos.

5.3 Testes comparativos

Dois testes foram desenvolvidos com objetivo de testar a hipótese 2, definida na seção 1.3, comparando o desempenho do método proposto com as abordagens correlatas descritas nas seções anteriores (5.1 e 5.2). A mesma estrutura e organização de simulação apresentada no capítulo 4 foi empregada para a realização dos testes comparativos.

Teste comparativo 1: Matriz de Avaliação com 71% de todas as transações registradas (esparsidade de 29%) durante treinamento, e mercado eletrônico composto por 8 Compradores e 32 Vendedores.

Teste comparativo 2: Matriz de Avaliação com 99% de todas as transações registradas (esparsidade de 1%) durante treinamento, e mercado eletrônico composto por 13 Compradores e 52 Vendedores.

O método proposto e os trabalhos correlatos apresentados no início deste capítulo, têm como preocupação em comum conferir um grau de confiabilidade adequado para o conjunto de avaliadores vizinhos selecionados pelo algoritmo tradicional de FCA. Apesar dos enfoques de seleção serem diferentes do método proposto – seleção de vizinhos honestos (Filtragem de Agrupamentos) [4], ou que não tenham sido selecionados por coincidência de avaliações (Peso de Relevância) [5] – consideramos importante estabelecer uma base de comparação em cenários de oscilação de comportamento. Todos os algoritmos compartilham configurações em comum, como o cálculo de similaridade entre vizinhos através do Coeficiente de Pearson, e a utilização do método de seleção de vizinhos por número de vizinhos (*best-n-neighbors*) [5, 14] configurado para 30 vizinhos.

Os métodos escolhidos para comparação com o método proposto foram:

- **Filtragem Colaborativa (FCA):** Compõe a técnica tradicional de Filtragem Colaborativa utilizada no *GroupLens* [1].
- **Filtragem de Agrupamentos (FCA2):** Mecanismo básico de Filtragem de Agrupamentos, proposto por Dellarocas [4], combina a técnica tradicional de FCA com o Algoritmo Divisor de Agrupamentos [48]. Para cada possível vendedor a ser recomendado, particiona o conjunto de vizinhos do comprador ativo em dois agrupamentos: avaliadores honestos e desonestos, e gera a estimativa de reputação com base no primeiro agrupamento. O particionamento pode ser feito em função da última avaliação ou da média das avaliações dos vizinhos sobre o vendedor, mas nenhuma função de distância específica é sugerida. No experimento optamos por utilizar uma função de distância Euclidiana baseada na última avaliação de cada vizinho sobre o vendedor.
- **Peso de Relevância (*Significance Weighting*) (FCA3):** É o método definido em [5], que aplica um fator de correção sobre o Coeficiente de Similaridade de cada vizinho calculado pelo algoritmo tradicional de FCA, com base na quantidade de itens em comum avaliados anteriormente com o usuário ativo. O fator de correção é:

$$\left\{ \begin{array}{ll} n/50 & \text{se } n < 50 \\ 1 & \text{se } n \geq 50 \end{array} \right.$$

onde n é a quantidade de avaliações em comum entre o avaliador e o comprador.

Observações importantes sobre o uso do Algoritmo de Filtragem de Agrupamentos no domínio de frequência:

Os experimentos realizados por Dellarocas visavam aferir a redução dos efeitos de avaliações desonestas na geração de recomendações, e utilizavam como principal parâmetro a proporção de avaliadores desonestos presentes no conjunto de vizinhos. Isto permitiu o teste de algumas estratégias de inserção de avaliadores desonestos no conjunto de vizinhos: 1) a escolha de avaliadores com valores de avaliação semelhantes aos do comprador para a maioria dos vendedores, exceto para o vendedor conivente (domínio de valor); 2) a inundação do sistema com avaliações desonestas (domínio de frequência), com o percentual de avaliadores desonestos controlado no experimento.

Nosso experimento explora situações de falsa vizinhança onde os avaliadores presentes no conjunto de vizinhos são tão inadequados quanto os avaliadores desonestos trabalhados por Dellarocas. Entretanto, estas situações são reproduzidas com a manipulação do comportamento de vendedores, e não dos avaliadores. A proporção de vendedores tipo Bad e Good cria as condições de oscilação de comportamento necessárias para a manifestação da falsa vizinhança. Em função dos avaliadores não serem variáveis controladas em nosso experimento, a efetividade do método proposto não foi avaliada em cenários de inundação de avaliações desonestas (*unfair ratings flooding*), e somente o algoritmo básico de Filtragem de Agrupamentos, voltado para a separação dos agrupamentos com base nos valores de avaliação (domínio de valor), foi implementado no experimento (a extensão do algoritmo voltada para separar os agrupamentos com base na frequência das avaliações não foi testada).

5.3.1 Análises estatísticas

O teste-t do tipo unilateral para amostras independentes foi aplicado utilizando nível de significância α em 5%. Cada amostra é composta pela distribuição das médias dos recursos financeiros nas rodadas de teste, que servem de base para medir o desempenho de cada método testado.

As hipóteses nula e alternativa foram definidas com base na hipótese 2 descrita no capítulo “Introdução”, seção 1.3:

H_0 – O desempenho do método proposto (designado como FCA1) é igual ao dos métodos comparados (FCA, FCA2 ou FCA3) em cenários de oscilação de comportamento dos vendedores entre as avaliações, denotada por:

$$H_0: \mu_{fca1} = \mu_{fcaN}$$

H_1 – O desempenho do método proposto (designado como FCA1) é melhor do que o dos métodos comparados (FCA, FCA2 ou FCA3) em cenários de oscilação de comportamento dos vendedores entre as avaliações:

$$H_1: \mu_{fca1} > \mu_{fcaN}$$

Conforme detalhado na seção “Atualização dos Recursos Financeiros” do capítulo Experimentos, o desempenho é medido com base na média dos recursos financeiros dos compradores. Os compradores apuram lucros em transações com vendedores de bom comportamento, e prejuízo com vendedores de mau comportamento. O melhor método em termos de desempenho é aquele que gerar o menor número de recomendações de vendedores de mau comportamento.

As hipóteses nula e de trabalho definidas nesta seção se aplicam aos testes comparativos 1 e 2 descritos nas seções 5.3.2 e 5.3.3.

5.3.2 Teste comparativo 1

Os mesmos cenários montados para os testes de recomendações (Tabela 4.3) do capítulo Experimentos foram usados nos testes comparativos. Para cada um dos seis cenários foram executadas quatro simulações independentes, cada qual correspondendo a um dos métodos testados (FCA, FCA1, FCA2 e FCA3), a fim de se garantir a aleatorização das amostras conforme explanado na seção “Organização da Simulação”.

Este teste usou a seguinte configuração (que reflete um cenário observado em sistemas de comércio consumidor para consumidor):

- 45 rodadas, sendo 25 de treinamento e 20 de teste (aproximadamente 29% de esparsidade da matriz de avaliação);

- 8 transações por rodada;
- 8 compradores ;
- 32 vendedores;
- total de 360 transações executadas.

5.3.2.1 Resultados

O resumo dos resultados é apresentado na Tabela 5.1.

Análise estatística dos melhores desempenhos						
	Cenário I	Cenário II	Cenário III	Cenário IV	Cenário V	Cenário VI
Bad	40%	50%	55%	60%	65%	75%
Good	50%	40%	35%	30%	25%	25%
Neutral	10%	10%	10%	10%	10%	0%
Melhor desempenho	FCA	FCA3	FCA3	FCA3	FCA3	FCA
2º Melhor desempenho	FCA3	FCA1	FCA1	FCA1	FCA1	FCA1
3º Melhor desempenho	FCA1	FCA	FCA	FCA	FCA	FCA3
Pior desempenho	FCA2	FCA2	FCA2	FCA2	FCA2	FCA2
Probabilidade de Significância do melhor em relação ao 2º melhor	0.4339	0.4856	0.3883	0.4697	0.4980	0.5832
Probabilidade de Significância do melhor em relação ao 3º melhor	0.3385	0.2237	0.3119	0.3339	0.1784	0.3940
Probabilidade de Significância do melhor em relação ao pior	0.0054	0.00002	0.0022	0.0001	0.0094	0.2749

Tabela 5.1 – Resumo dos resultados dos testes

A Tabela 5.1 indica os melhores desempenhos entre as abordagens testadas por cenário de teste. A probabilidade de significância indica se o desempenho do melhor método é significativamente maior, em termos estatísticos, em relação aos demais. Quanto maior a probabilidade de significância menos significativa é a diferença.

Através da análise da tabela verifica-se que FCA3 apresenta o melhor desempenho em 4 dos 6 cenários testados, enquanto FCA em 2 cenários. O desempenho do método proposto FCA1 se mostra estável sendo o segundo melhor em 5 dos 6 cenários testados. Entretanto, observa-se que a diferença de desempenho entre estes 3 métodos não chega a ser significativa em nenhum momento. Já com relação a FCA2, que apresenta o pior desempenho em todos os cenários testados, as diferenças dos 3 métodos são significativas em quase todos os cenários, com exceção do cenário VII.

De uma forma geral, FCA3 e FCA1 parecem se beneficiar com o aumento da proporção de vendedores tipo Bad. Após o cenário I o desempenho deles passa a ser melhor do que FCA, e permanece assim até o cenário V. Isto ocorre em função da maior degradação de desempenho sofrida por FCA entre os cenários I e II, conforme observado na Tabela 5.2:

Média geral dos recursos financeiros						
	Cenário I	Cenário II	Cenário III	Cenário IV	Cenário V	Cenário VI
Bad	40%	50%	55%	60%	65%	75%
Good	50%	40%	35%	30%	25%	25%
Neutral	10%	10%	10%	10%	10%	0%
FCA	5067.78	5044.57	5028.12	5020.15	4989.54	4980.59
FCA1	5063.18	5049.64	5029.15	5022.64	4998.07	4972.43
FCA2	5042.14	5021.37	5016.96	4999.01	4977.92	4972.01
FCA3	5065.95	5049.92	5031.03	5023.20	4998.03	4976.65

Tabela 5.2 – Média geral dos recursos financeiros por cenário

A partir do cenário IV há uma degradação mais acentuada dos desempenhos de FCA3 e FCA1, que acaba por aproximá-los à FCA2 e por ficar abaixo do desempenho de FCA no último cenário.

Os gráficos das figuras 5.1 a 5.6 demonstram a comparação do desempenho financeiro dos compradores por método avaliado e cenário de teste. Da mesma forma que nos testes de geração de recomendação (seção 4.3), descrevem a média dos recursos financeiros dos compradores ao longo das rodadas de teste. A diferença é que são definidas e comparadas 4 séries, cada qual associada a um dos métodos testados.

Assim como nos testes de recomendação, foi observado que a curva de declínio tende a se alongar à medida que aumenta a proporção de vendedores tipo Bad, iniciando na rodada 42 quando há 40% de tipo Bad (Figura 5.1), e na rodada 35 quando há 75% de tipo Bad (Figura 5.6).

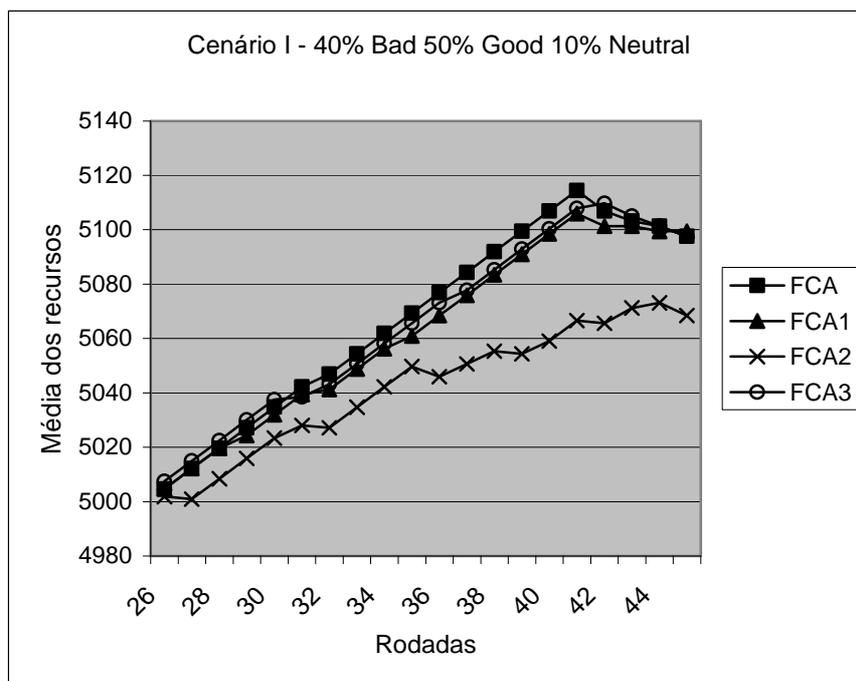


Figura 5.1 – Comparação do desempenho financeiro entre os métodos testados no cenário I

O gráfico da Figura 5.1 demonstra que na rodada 42 o método FCA alcança o ápice de desempenho financeiro do teste efetuado, com a marca aproximada de 5150 unidades monetárias. A partir desta rodada se inicia o declínio do desempenho médio financeiro dos compradores que utilizam os métodos FCA, FCA1 e FCA3, sendo o declive um pouco mais acentuado para o método FCA1, e equivalente entre os métodos FCA e FCA3. Também é observado o baixo desempenho financeiro do grupo de compradores que utilizam o método FCA2, que desde as rodadas iniciais se mostra bem menor do que os demais métodos. Por último, é verificado que nenhum dos métodos testados fez com que o desempenho médio financeiro atingisse um patamar inferior ao inicial (5000 unidades monetárias) ao final do teste.

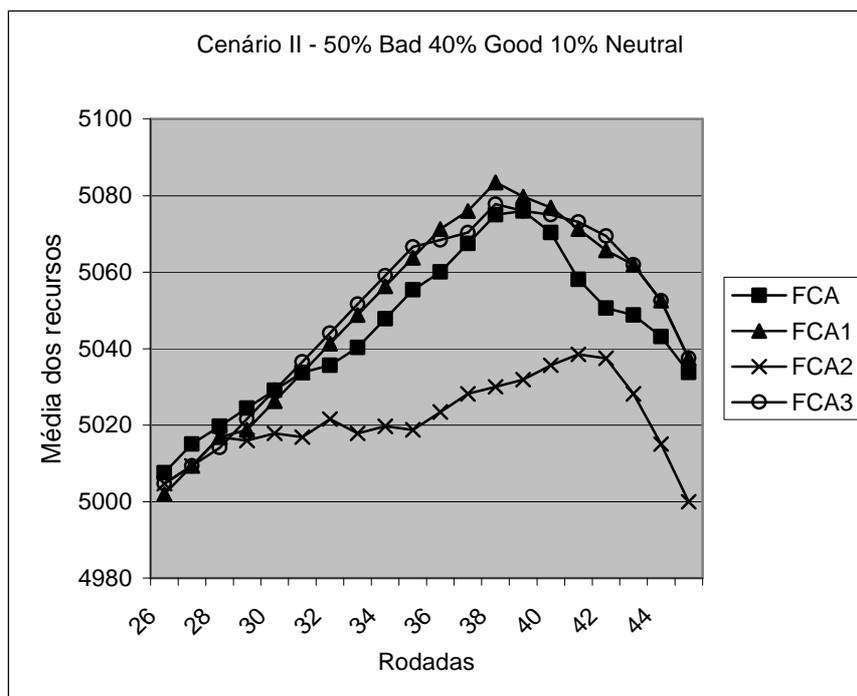


Figura 5.2 – Comparação do desempenho financeiro entre os métodos testados no cenário II

O gráfico da Figura 5.2 demonstra que na rodada 39 o método FCA1, e não mais FCA, alcança o ápice de desempenho financeiro do teste efetuado, com a marca aproximada de 5080 unidades monetárias. A partir desta rodada se inicia o declínio do desempenho médio financeiro dos compradores que utilizam os métodos FCA, FCA1 e FCA3, sendo o declive um pouco mais acentuado para o método FCA, e equivalente entre os métodos FCA1 e FCA3. A diferença do desempenho financeiro do grupo de compradores que utilizam o método FCA2 se acentuou em relação ao cenário anterior testado, fazendo com que o desempenho médio financeiro deste grupo atingisse o patamar inicial de 5000 unidades monetárias ao final do teste.

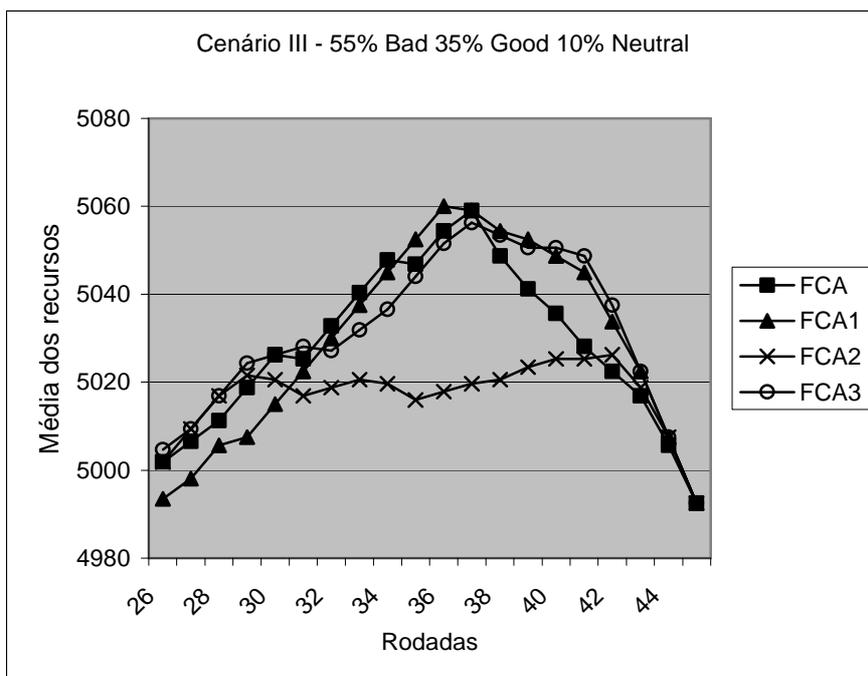


Figura 5.3 – Comparação do desempenho financeiro entre os métodos testados no cenário III

O gráfico da Figura 5.3 demonstra que na rodada 37 o método FCA1 alcança o ápice de desempenho financeiro do teste efetuado, com a marca aproximada de 5060 unidades monetárias. O ápice e declínio do desempenho de FCA1 ocorre uma rodada antes de FCA e FCA3. Dos métodos FCA, FCA1 e FCA3, o maior declínio de desempenho é observado para FCA, e o menor para FCA3. Ao contrário dos cenários anteriores, o desempenho médio financeiro de FCA2 se mantém por algumas rodadas (26 a 30) superior ao de FCA e FCA1. A partir da rodada 44 o desempenho dos 4 métodos testados alcança o mesmo patamar. Deste momento em diante não se observa diferença do desempenho financeiro para o método FCA2. Ao final do teste, o desempenho médio financeiro de todos os grupos de compradores atinge um patamar inferior ao inicialmente definido (5000 unidades monetárias), o que indica maior degradação dos métodos testados perante o aumento da proporção de vendedores do tipo *Bad*.

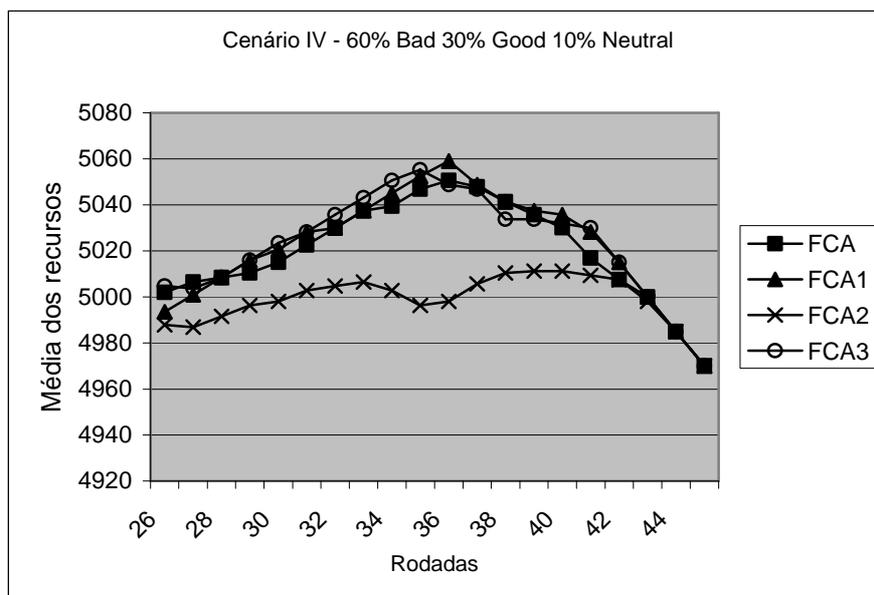


Figura 5.4 – Comparação do desempenho financeiro entre os métodos testados no cenário IV

O gráfico da Figura 5.4 demonstra que o ápice e declínio do desempenho de FCA3 ocorre uma rodada antes de FCA e FCA1, na rodada 36. O método FCA1 mantém o comportamento do cenário anterior alcançando o ápice de desempenho financeiro do teste na rodada 37, com a marca aproximada de 5060 unidades monetárias. A partir da rodada 37 o declínio de desempenho é similar para FCA, FCA1 e FCA3. Da mesma forma que no cenário anterior, a partir da rodada 44 o desempenho dos 4 métodos testados alcança o mesmo patamar. Ao final do teste, o desempenho médio financeiro de todos os grupos de compradores é não só inferior ao inicialmente definido, como também inferior ao do cenário anterior (em torno de 4970 contra 4990 unidades monetárias do cenário anterior).

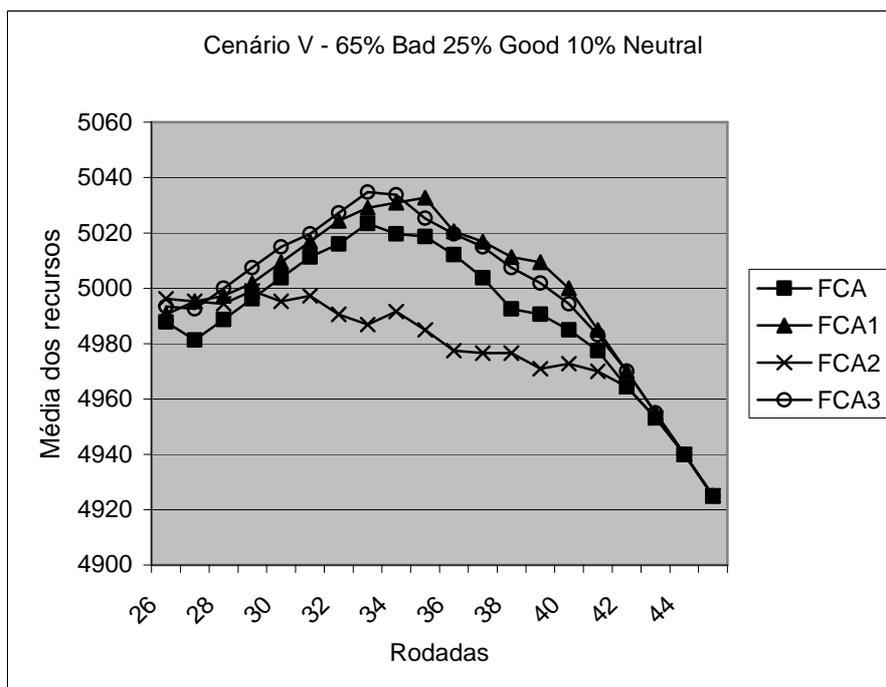


Figura 5.5 – Comparação do desempenho financeiro entre os métodos testados no cenário V

O gráfico da Figura 5.5 demonstra que o ápice e declínio do desempenho de FCA3 e FCA ocorre uma rodada antes de FCA1, na rodada 34. O método FCA3 alcança na rodada 34 o ápice de desempenho financeiro do teste efetuado, com a marca aproximada de 5030 unidades monetárias. O desempenho de FCA se mantém inferior ao de FCA1 e FCA3 ao longo de todas as rodadas. A partir da rodada 44 o desempenho dos 4 métodos testados alcança o mesmo patamar. A tendência de degradação do desempenho dos métodos, observada no cenário anterior, é mantida com a apuração de desempenho médio financeiro em torno de 4920 unidades monetárias, contra 4970 do cenário anterior.

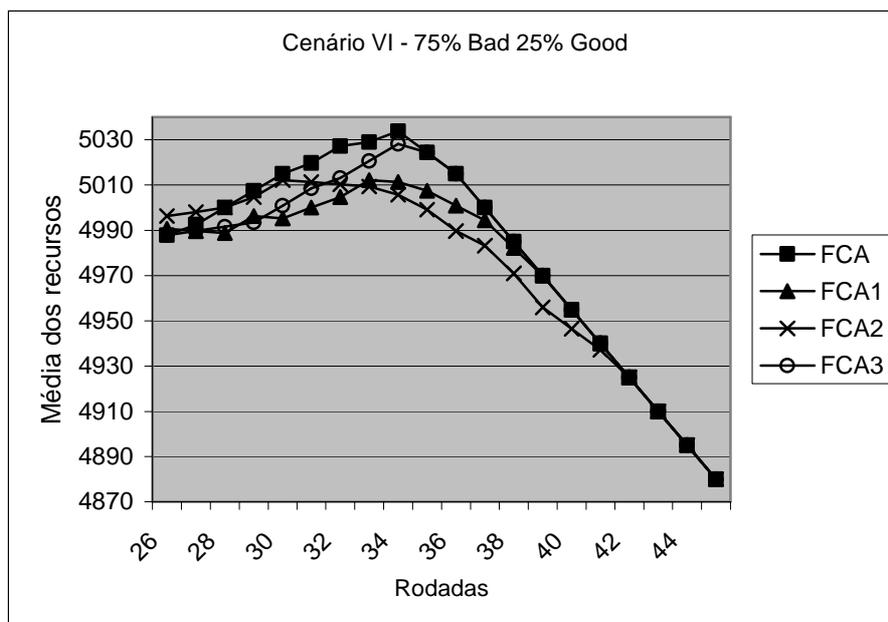


Figura 5.6 – Comparação do desempenho financeiro entre os métodos testados no cenário VI

O gráfico da Figura 5.6 demonstra que o ápice e declínio do desempenho de FCA1 ocorre uma rodada antes de FCA e FCA3, na rodada 34. O método FCA volta a alcançar o ápice de desempenho financeiro do teste efetuado, com a marca aproximada de 5030 unidades monetárias na rodada 35. O desempenho de FCA1 se mantém inferior ao de FCA e FCA3 ao longo de todas as rodadas. Ao contrário do observado nos testes anteriores, o desempenho de FCA, FCA1 e FCA3 se mantém mais próximo de FCA2 ao longo de todas as rodadas. Já a partir da rodada 42 o desempenho dos 4 métodos testados alcança o mesmo patamar. A tendência de degradação dos métodos é mantida com a apuração de desempenho médio financeiro em torno de 4880 unidades monetárias, contra 4920 do cenário anterior.

Na maioria dos cenários foram observadas oscilações nas curvas do método FCA2, que podem ser explicadas com a infiltração de vendedores tipo Bad desde as rodadas iniciais. Os gráficos das figuras 5.7 e 5.8 demonstram isto:

FCA2 - Cenário II - Distribuição de Recomendações

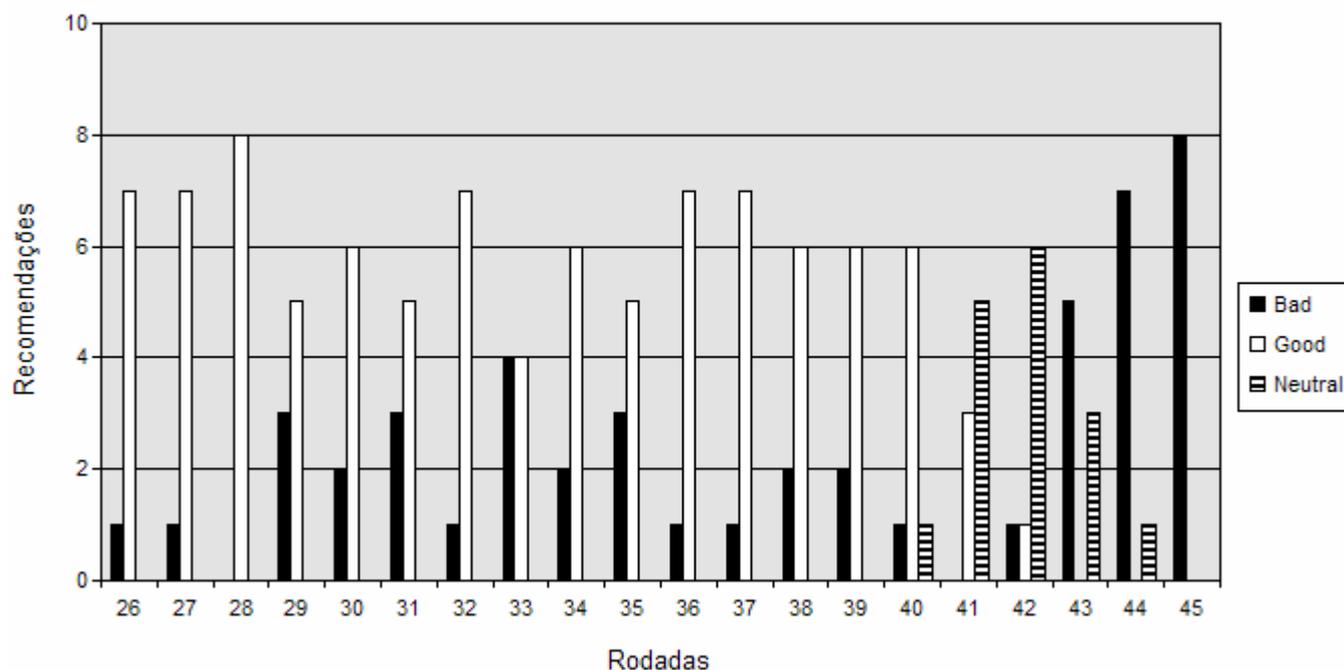


Figura 5.7 – Distribuição de recomendações do método FCA2 no cenário II

FCA2 - Cenário III - Distribuição de Recomendações

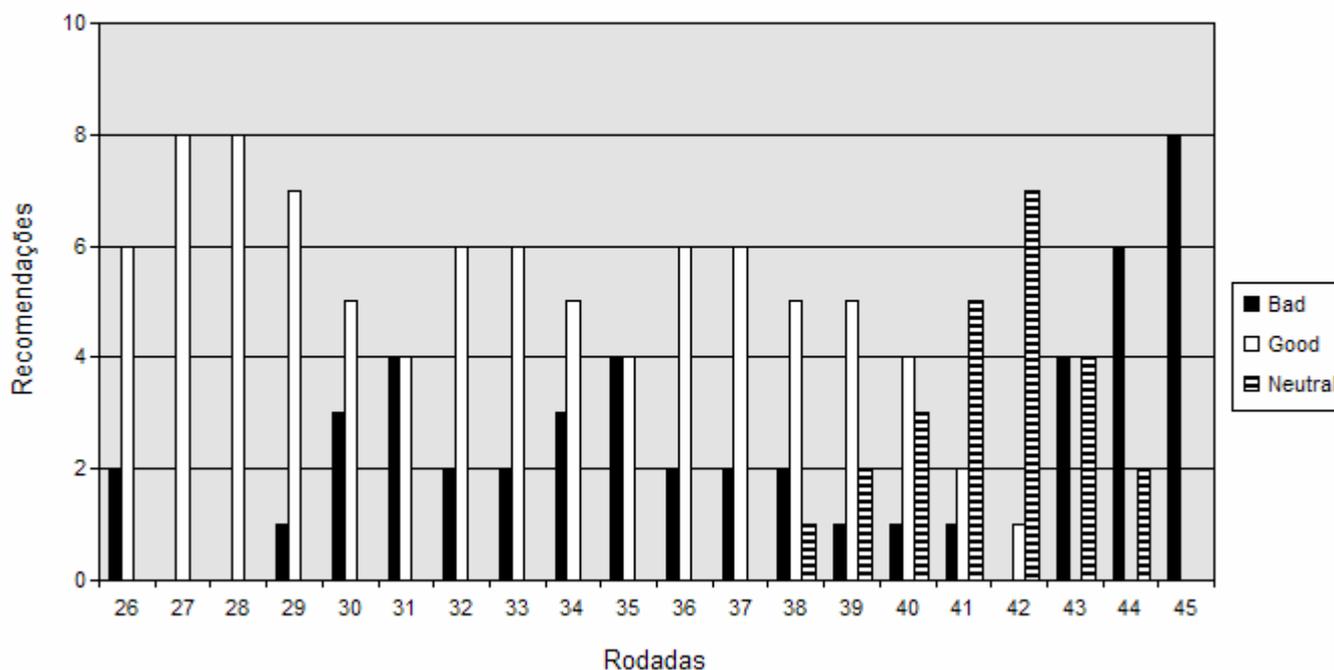


Figura 5.8 – Distribuição de recomendações do método FCA2 no cenário III

Nos gráficos das figuras 5.7 e 5.8, o eixo y identifica a quantidade de recomendações de vendedores com perfil *Bad*, *Good*, e *Neutrals* geradas pelo método FCA2. Nos dois cenários apresentados as recomendações do tipo *Bad* são geradas desde as rodadas iniciais, antes mesmo das recomendações do tipo *Neutral*, que começam a ser geradas somente nas rodadas 40 e 38 respectivamente.

Análises preliminares levantam a suposição de que o melhor desempenho geral de FCA3 pode ser atribuído à baixa quantidade de avaliações em comum entre dois compradores (32, que é a quantidade de vendedores configurada para o experimento), que acaba por diminuir a probabilidade dos compradores se depararem com situações de oscilação de comportamento². A aplicação do fator de correção também pode ter influenciado na minimização do impacto de avaliações de falsos bons vizinhos à medida que é aplicado um fator redutor de 32/50 (.64) sobre o coeficiente de similaridade apurado para cada vizinho. Em função destas considerações optou-se por realizar um segundo teste comparativo, com maior quantidade de vendedores simulados, de modo a verificar se o aumento da quantidade de vendedores teria um efeito contrário ao obtido no primeiro teste comparativo, culminando com a queda do desempenho de FCA3 em relações aos demais métodos.

Com relação ao baixo desempenho de FCA2, verificou-se que, na média, cada vendedor foi avaliado durante as rodadas de treino por cerca de 5.75 compradores, o que corresponde a 71% do total máximo de 8 compradores configurados na simulação. Com isso o conjunto de vizinhos disponível para cada recomendação se tornou muito pequeno, e com maiores chances de todos os elementos terem sido influenciados pela oscilação de comportamento dos vendedores, o que pode ter levado a uma divisão de agrupamentos pouco efetiva. Optou-se por no novo teste comparativo aumentar o tamanho do conjunto de vizinhos para 13 e verificar se o algoritmo de divisão de agrupamentos (análise de dissimilaridade) se beneficia deste aumento, retirando de forma mais efetiva os falsos vizinhos do grupo de vizinhos honesto de cada recomendação.

5.3.2.2 Conclusões

De acordo os resultados apresentados na Tabela 5.1 da seção 5.3.2.1, concluímos que o teste não rejeita a hipótese nula H_0 para os métodos FCA3 e FCA (ao nível de significância de 5%), com o desempenho de FCA1 tendendo a ser igual ao deles. Já com relação ao método FCA2, o teste rejeita a hipótese H_0 em 6 dos 7 cenários testados, indicando que FCA1 tende a ser melhor do que FCA2.

² . Conforme discutido na seção 5.2 - Peso de Relevância, quanto maior a quantidade de avaliações maior a probabilidade de se deparar com usuários com comportamento oscilante, e maior a chance de se selecionar falsos bons vizinhos.

Em vista das condições deste teste ainda não serem as ideais para efetivamente se avaliar FCA2 e FCA3, os testes de hipóteses aqui registrados foram reaplicados em um outro teste comparativo descrito a seguir, para se confirmar a manutenção dos resultados obtidos.

5.3.3 Teste comparativo 2

A primeira mudança em relação aos testes anteriores foi o aumento da quantidade de compradores para 13 e da duração do treinamento para 81 rodadas, fazendo com que ao seu término tivessem sido realizadas cerca de 99% das possíveis transações. Desta forma, seria garantida durante as rodadas de teste a maximização do tamanho do conjunto de vizinhos disponível em todas as recomendações (cerca de 13 avaliadores), e seria possível avaliar se o desempenho de FCA2 melhora nestas condições. A segunda mudança foi na quantidade de vendedores que subiu para 52. Desta forma, haveria o aumento da exposição dos compradores a situações de oscilação de comportamento dos vendedores, e a quantidade de avaliações em comum passaria para um patamar superior ao limite usado pelo fator de redução ($n/50$), de modo a ser possível verificar se estes fatores influenciam negativamente no desempenho de FCA3.

Os mesmos cenários montados para os testes de recomendações (Tabela 4.3) do capítulo Experimentos foram usados nos testes comparativos. Para cada um dos seis cenários foram executadas quatro simulações independentes, cada qual correspondendo a um dos quatro métodos testados (FCA, FCA1, FCA2 e FCA3), a fim de se garantir a aleatorização das amostras conforme explicado na seção “Organização da Simulação”.

Este teste usou a seguinte configuração:

- 114 rodadas, sendo 81 de treinamento e 34 de teste (aproximadamente 1% de esparsidade da matriz de avaliação);
- 13 transações por rodada;
- 13 compradores ;
- 52 vendedores;
- em torno de 1100 transações executadas.

5.3.3.1 Resultados

O resumo dos resultados é apresentado na Tabela 5.3.

Análise estatística dos melhores desempenhos						
	Cenário I	Cenário II	Cenário III	Cenário IV	Cenário V	Cenário VI
Bad	40%	50%	55%	60%	65%	75%
Good	50%	40%	35%	30%	25%	25%
Neutral	10%	10%	10%	10%	10%	0%
Melhor desempenho	FCA	FCA2	FCA2	FCA2	FCA2	FCA2
2º Melhor desempenho	FCA3	FCA1	FCA1	FCA1	FCA1	FCA
3º Melhor desempenho	FCA1	FCA	FCA	FCA	FCA	FCA3
Pior desempenho	FCA2	FCA3	FCA3	FCA3	FCA3	FCA1
Probabilidade de Significância do melhor em relação ao 2º melhor	0.3181	0.4926	0.3641	0.3794	0.4818	0.4933
Probabilidade de Significância do melhor em relação ao 3º melhor	0.3371	0.4742	0.3291	0.3519	0.4263	0.4771
Probabilidade de Significância do melhor em relação ao pior	0.3586	0.2797	0.2761	0.3170	0.4089	0.4666

Tabela 5.3 – Resumo dos resultados dos testes

Através da análise da tabela verifica-se que FCA2 apresenta o melhor desempenho em 5 dos 6 cenários testados, enquanto FCA em 1 cenário. O desempenho do método proposto FCA1 se mostra estável sendo o segundo melhor em 4 dos 6 cenários testados. Entretanto, observa-se que a diferença de desempenho entre os 4 métodos não chega a ser significativa em nenhum momento. No âmbito geral FCA3 apresentou o pior desempenho, sendo o pior em 4 dos 6 cenários testados.

O cenário I apresentou a mesma ordem de desempenho do primeiro teste comparativo (FCA, FCA3, FCA1, FCA2). Entretanto, ao contrário do primeiro teste, a diferença de desempenho dos demais métodos para FCA2 não foi significativa.

De uma forma geral, FCA2 e FCA1 se beneficiam com o aumento da proporção de vendedores tipo Bad. No caso de FCA2, seu desempenho se mantém melhor até o último cenário, entretanto, após a sua ascensão ao primeiro posto no cenário II, a diferença para os demais métodos cai gradativamente nos cenários subsequentes. Já FCA1 se mantém em segundo lugar até o cenário V, mas por apresentar uma queda mais acentuada, acaba no último cenário tendo o pior desempenho dos 4 métodos.

Os gráficos das figuras 5.9 a 5.14 mostram a comparação do desempenho financeiro dos compradores por método avaliado e cenário de teste.

Para facilitar a visualização dos gráficos são mostradas apenas as rodadas onde as diferenças dos recursos financeiros são mais acentuadas. À medida que a proporção de Bad aumenta nos cenários, as diferenças de desempenho se acentuam mais cedo, da rodada 107 no cenário I para a rodada 94 no cenário VI.

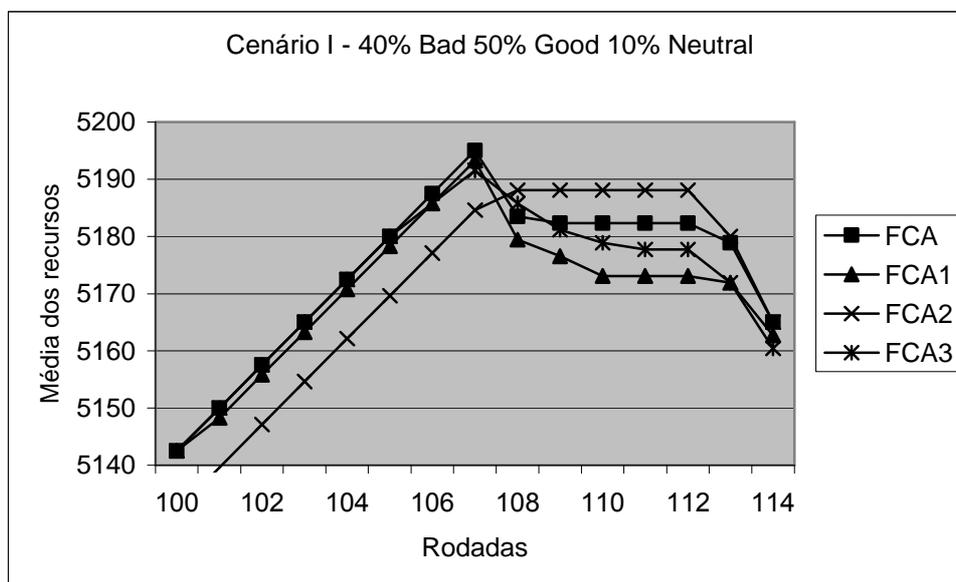


Figura 5.9 – Comparação do desempenho financeiro dos métodos no cenário I

No cenário I (Figura 5.9) verificou-se que o desempenho de FCA2 se manteve crescente, mas por muitas rodadas abaixo dos demais, por recomendar alguns vendedores com perfil Bad desde o início da fase de teste. Entre as rodadas 108 e 113 seu desempenho foi superior aos demais sem registrar queda em vista da seleção de vendedores neutrais, entretanto isto não foi suficiente para tirar FCA2 do pior desempenho geral. Nestas rodadas, FCA1 apresentou o maior declínio de recursos, mas por ter se mantido equivalente aos demais na maior parte do tempo, acabou não tendo o pior desempenho geral. Entre as rodadas 110 e 112 observou-se que todos os métodos testados proporcionaram desempenho financeiro estável para os compradores, em razão da substituição de recomendações do tipo *Bad* pelo tipo *Neutral*. Esta tendência se reverteu em declínio a partir da rodada 113.

Entre os cenários II e V (Figuras 5.10 a 5.14) os recursos financeiros dos compradores que utilizavam FCA1 e FCA2 se mantiveram equivalentes aos demais durante as rodadas iniciais, e tiveram

menor queda durante as rodadas de declínio, sendo que FCA2 apresentou a menor queda de todos. Em função disto o desempenho geral destes métodos foi melhor nestes cenários.

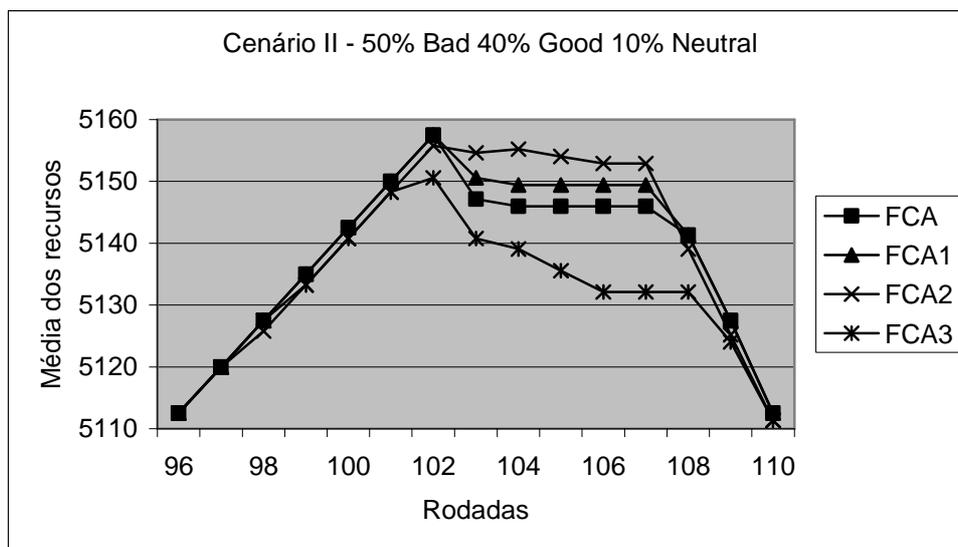


Figura 5.10 – Comparação do desempenho financeiro dos métodos no cenário II

Na Figura 5.10, após queda do desempenho financeiro, os métodos FCA, FCA1 e FCA2 atingem patamar de estabilidade entre as rodadas 103 e 107. Dentre eles, FCA atinge o menor patamar, em torno de 5148 unidades monetárias. A maior queda de desempenho do teste ocorre com FCA3, cuja estabilidade só ocorre entre as rodadas 106 e 108, se mantendo no patamar de 5130 unidades monetárias.

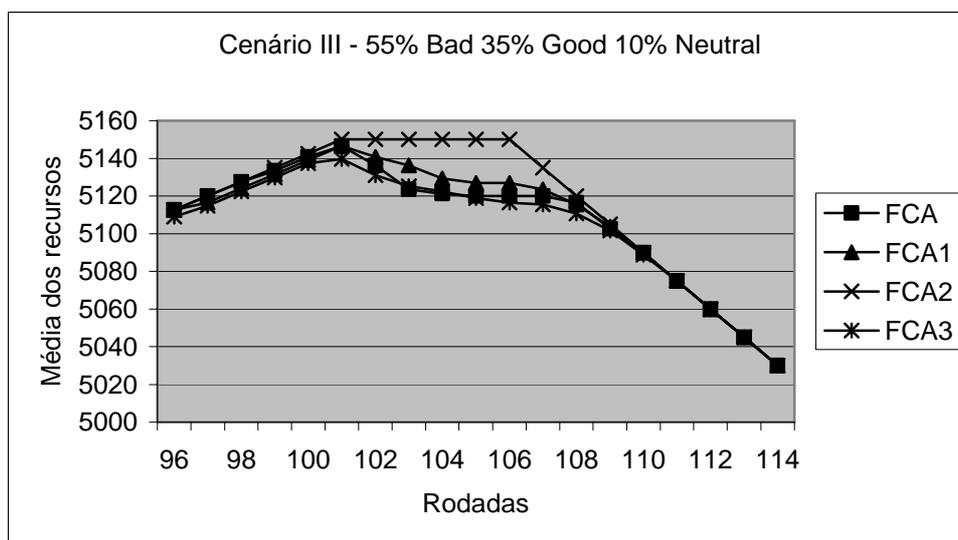


Figura 5.11 – Comparação do desempenho financeiro dos métodos no cenário III

Nas Figuras 5.11 e 5.12, o período de estabilidade de FCA2 ocorre sem ser precedido por queda de desempenho, o que confere o melhor desempenho do teste. A queda é equivalente para FCA e FCA3, e ocorre com menor intensidade para FCA1, que se mantém em um patamar um pouco superior.

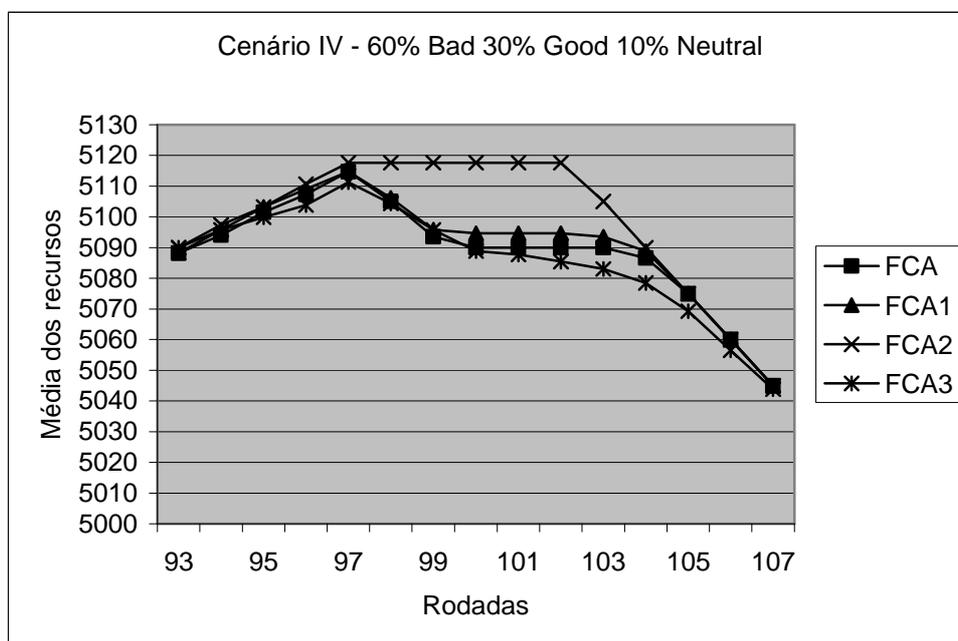


Figura 5.12 – Comparação do desempenho financeiro dos métodos no cenário IV

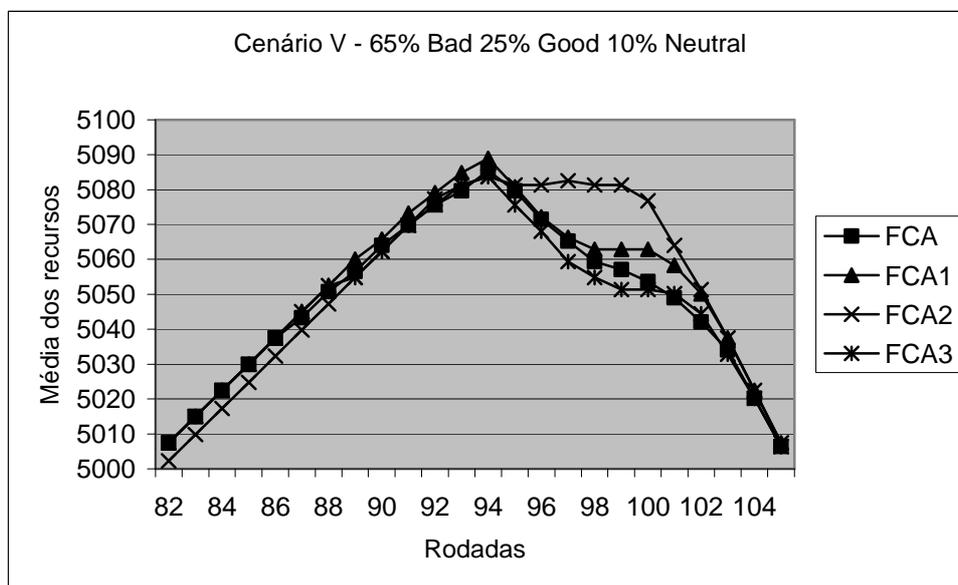


Figura 5.13 – Comparação do desempenho financeiro dos métodos no cenário V

Na Figura 5.13, o período de estabilidade de FCA2 ocorre precedido por uma queda de desempenho menos acentuada em relação aos demais métodos, o que confere o melhor desempenho do teste. A queda de desempenho é equivalente para FCA e FCA3, e ocorre com menor intensidade para FCA1, que se mantém em um patamar equivalente ao de FCA2 a partir da rodada 102.

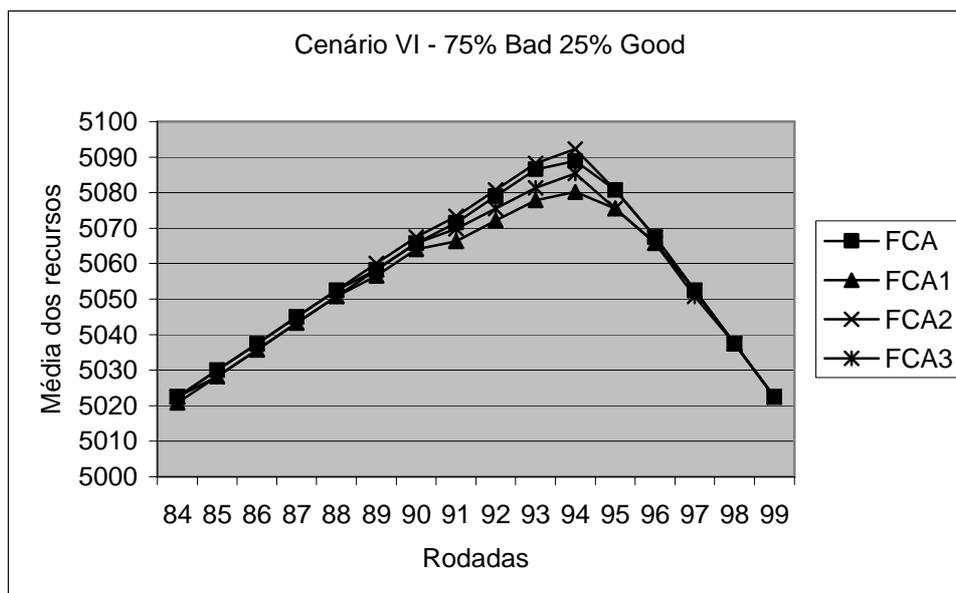


Figura 5.14 – Comparação do desempenho financeiro dos métodos no cenário VI

Na Figura 5.14 não ocorre o período de estabilidade de desempenho financeiro observado nos cenários de teste anteriores. O ápice de desempenho continua sendo dos compradores que utilizam o método FCA2, que na rodada 94 alcança o patamar de 5090 unidades monetárias. O menor ápice ocorre com FCA1 que na rodada 94 alcança a marca de 5080 unidades monetárias.

Os diferenciais de desempenho de FCA2 e FCA1 em relação aos demais podem ser explicados com os gráficos a seguir.

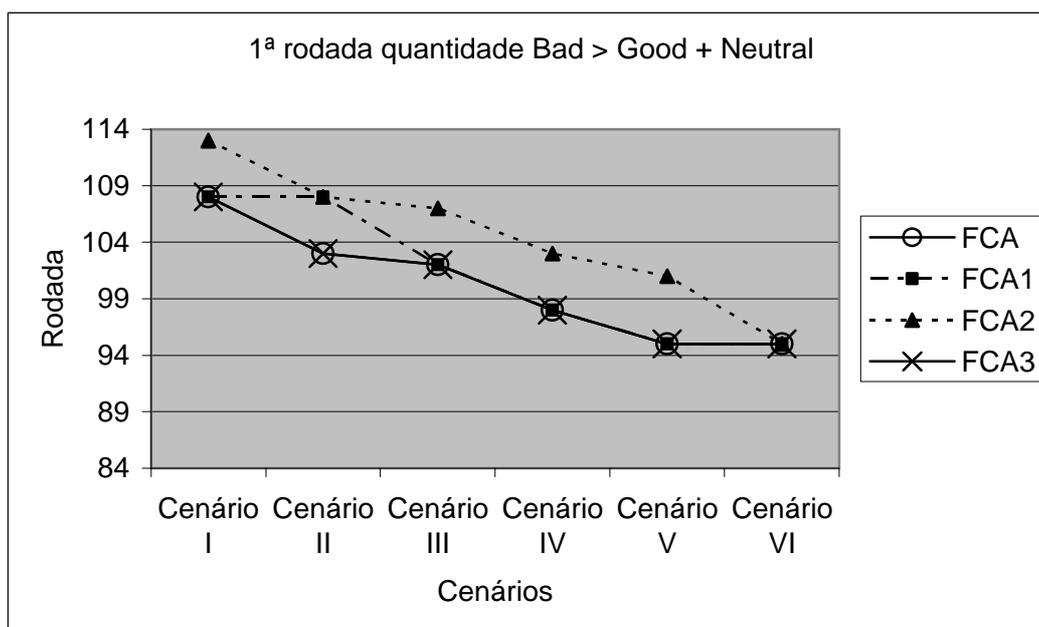


Figura 5.15 – 1ª Rodada onde a quantidade de Bad é maior que a quantidade de Good + Neutral

O gráfico da Figura 5.15 mostra a primeira rodada onde a quantidade de Bad é maior do que a quantidade de Good e Neutral recomendados. Isto ocorre com FCA2 somente em rodadas posteriores às dos outros métodos porque prioriza recomendações de Good e Neutral em relação às recomendações de Bad, enquanto os demais também priorizam Good mas intercalam Neutral com Bad (Figura 5.16 a Figura 5.19).

O patamar de recomendação de FCA1 equivale ao de FCA e FCA3 em quase todos os casos, entretanto seu desempenho permanece melhor em razão da proporção de Bad recomendados ser menor durante os cenários II a V (Figura 5.20). Conclui-se que apesar da sua distribuição de recomendações não ser tão equilibrada quanto a de FCA2, ainda assim é menos vulnerável a infiltrações de Bad nas rodadas iniciais do que FCA e FCA3.

O patamar mínimo de recomendação (rodada 95) passou a ser o mesmo para todos os métodos testados, inclusive FCA2, no cenário VI.

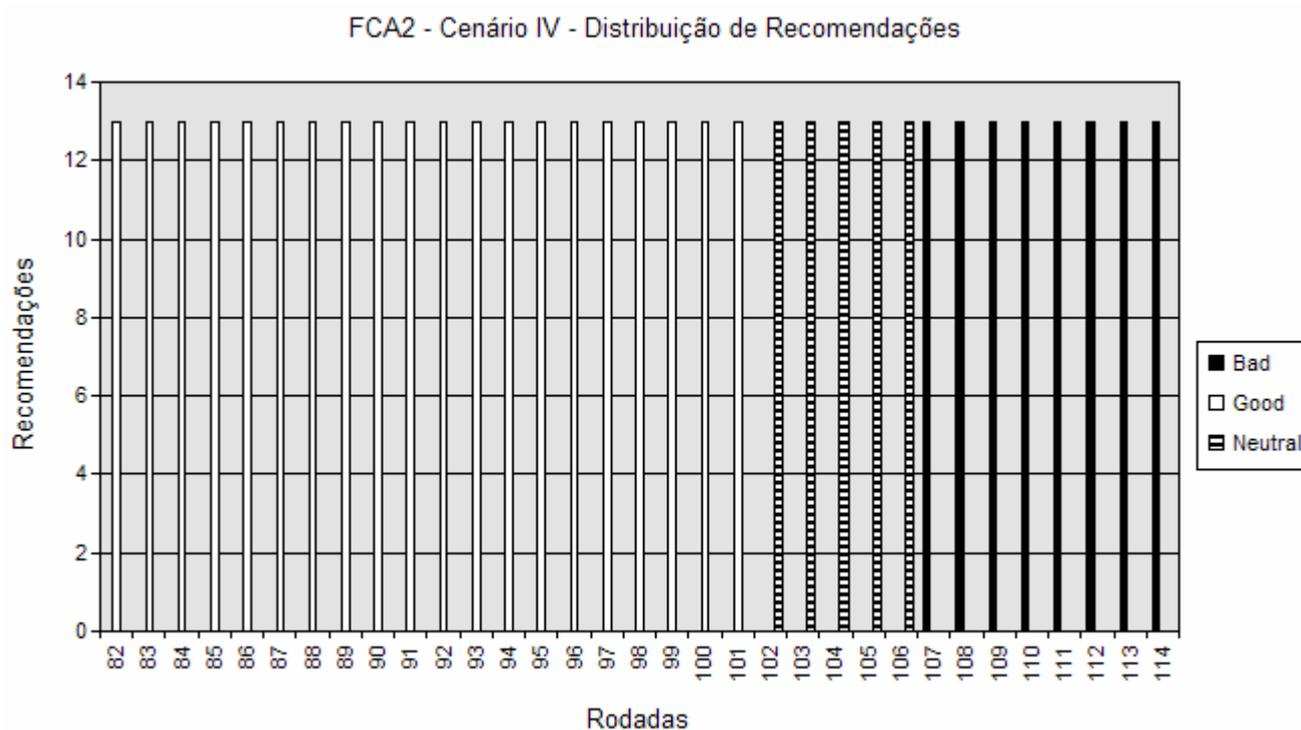


Figura 5.16 – Distribuição de recomendações de FCA2 com priorização de Good e Neutral em relação a Bads

FCA1 - Cenário IV - Distribuição de Recomendações

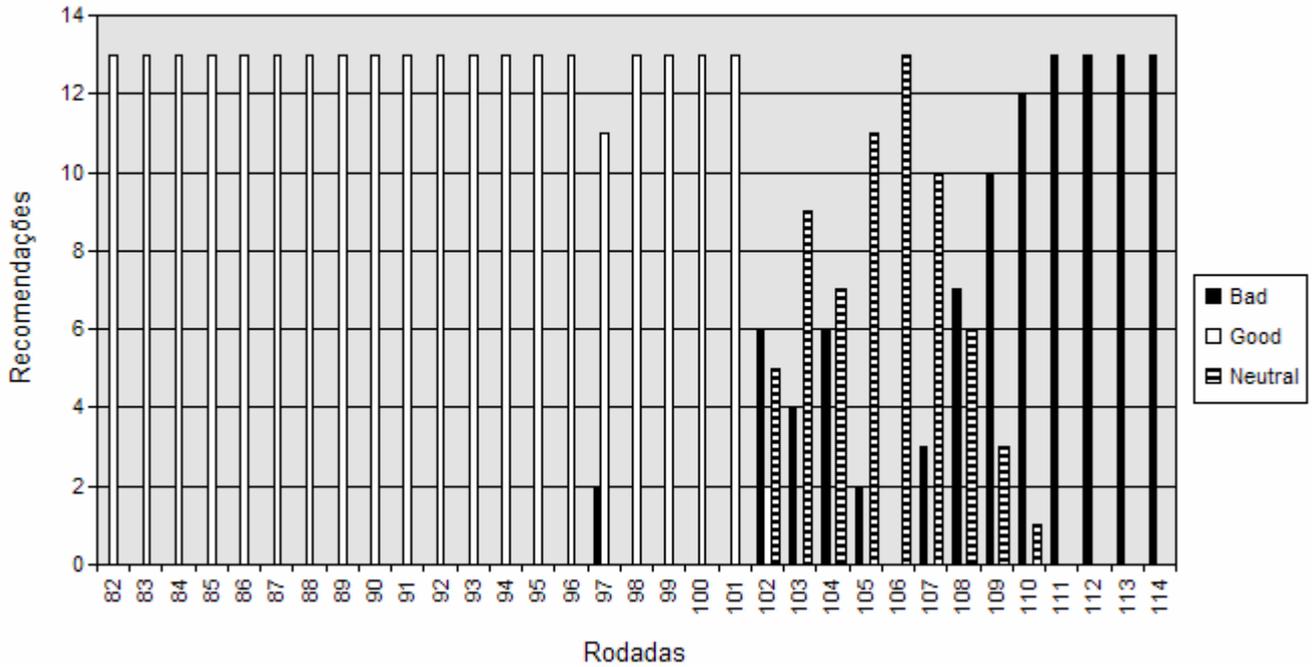


Figura 5.17 – Distribuição de recomendações de FCA1 com intercalação de Neutral e Bads nas recomendações intermediárias

FCA - Cenário IV - Distribuição de Recomendações

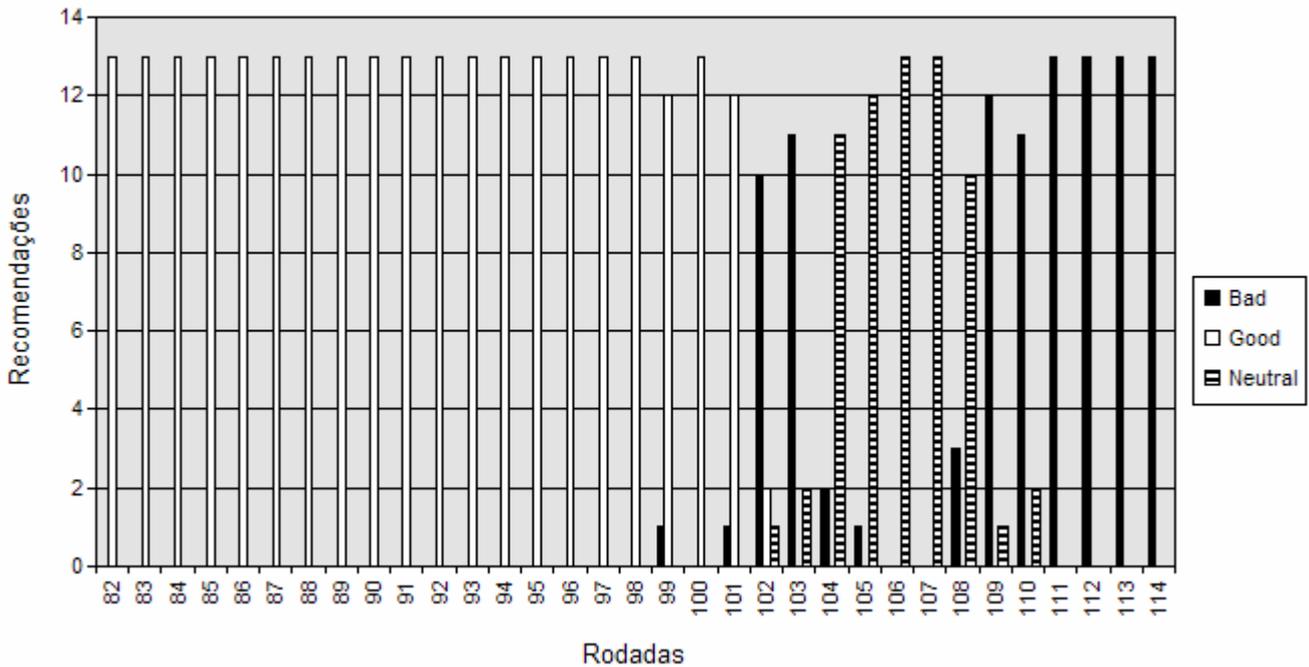


Figura 5.18 – Distribuição de recomendações de FCA com intercalação de Neutral e Bads nas recomendações intermediárias

FCA3 - Cenário IV - Distribuição de Recomendações

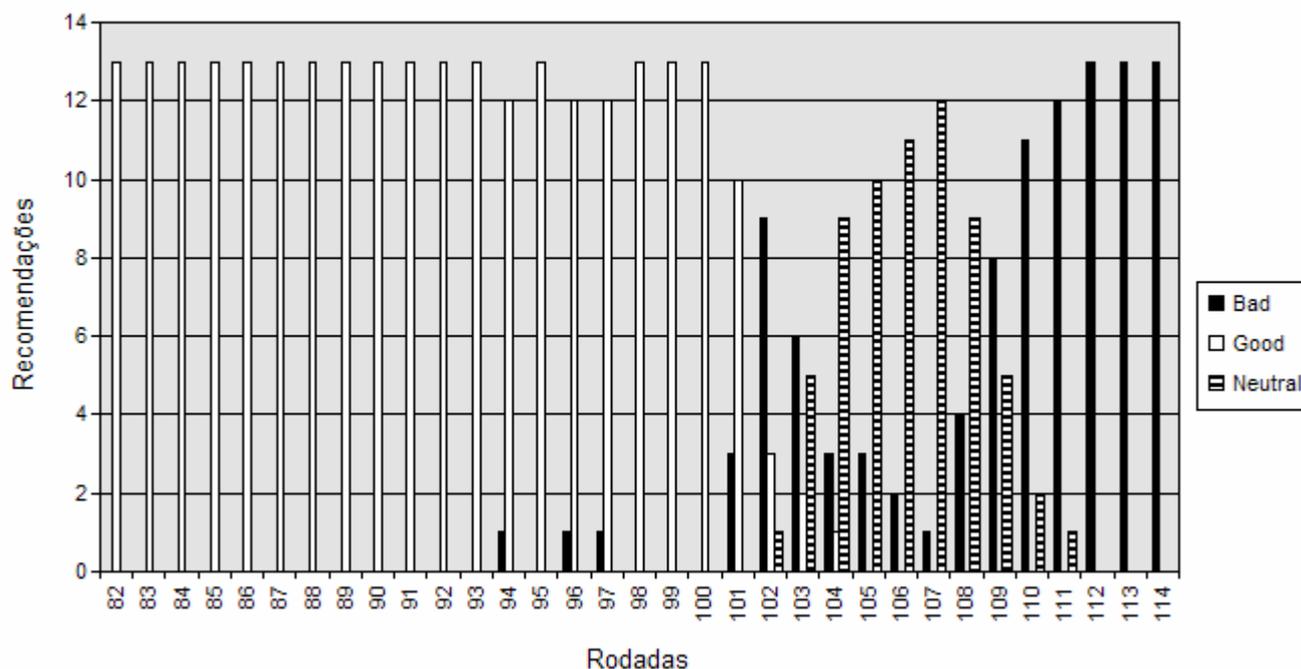


Figura 5.19 – Distribuição de recomendações de FCA3 com intercalação de Neutral e Bads nas recomendações intermediárias

Nos gráficos das Figuras 5.16 a 5.19 percebe-se que a partir da rodada 102 já não são geradas recomendações de *Goods*. Observa-se também que os métodos FCA2 e FCA1 geram menor quantidade de recomendações *Bad* até esta rodada, 0 e 8 respectivamente, contra 12 e 15 para FCA e FCA3.

Quantidade de vendedores Bad recomendados

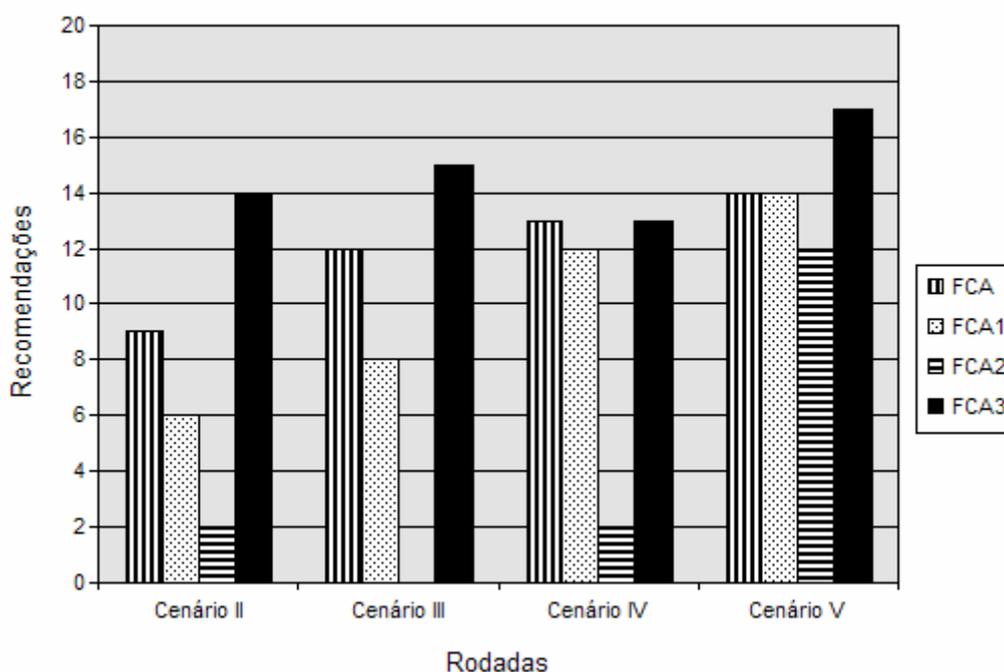


Figura 5.20 – Menor quantidade de recomendações de vendedores Bad de FCA1 e FCA2 em relação a FCA e FCA3

No gráfico da Figura 5.20 se observa que, em todos os cenários testados, a quantidade de vendedores *Bad* recomendada pelos métodos FCA1 e FCA2 foi menor do que a dos métodos FCA e FCA3.

5.3.3.2 Conclusões

De acordo os resultados apresentados na Tabela 5.3 da seção 5.3.3.1, concluímos que o teste não rejeita a hipótese nula H_0 para nenhum dos métodos comparados com o método proposto (ao nível de significância de 5%), com o desempenho de FCA1 tendendo a ser igual ao deles.

No teste comparativo anterior (seção 5.3.2), o método proposto apresentou desempenho significativamente melhor do que FCA2, entretanto esta diferença parece ter sido causada mais em função da limitada quantidade de avaliadores vizinhos presente no teste anterior, do que por um desempenho efetivamente melhor do método proposto. Neste teste constatou-se a melhora do desempenho de FCA2 com a maximização do tamanho do conjunto de vizinhos. De acordo com a Tabela 5.4, ao final da fase de treinamento, dos 52 vendedores existentes, 49 foram avaliados por todos os 13 compradores existentes, e 3 vendedores foram avaliados por 12 compradores. Isto significa que para qualquer recomendação da fase de teste, o conjunto de vizinhos teve no mínimo 12 avaliadores disponíveis, contra os 5.75 do teste comparativo anterior.

Quanto ao método FCA3, foi constatada a queda de desempenho levantada como suposição no primeiro teste comparativo. À medida que se aumentou a exposição dos compradores a situações de oscilação de comportamento dos vendedores, e o fator de correção passou a não mais agir como redutor do coeficiente de similaridade (Tabela 5.5), verificou-se queda acentuada do desempenho de FCA3 em comparação aos demais métodos.

Vendedor	Quantidade de compradores	Vendedor	Quantidade de compradores
Vendedor0	13	Vendedor26	13
Vendedor1	13	Vendedor27	13
Vendedor2	13	Vendedor28	13
Vendedor3	13	Vendedor29	13
Vendedor4	13	Vendedor30	13
Vendedor5	13	Vendedor31	13
Vendedor6	13	Vendedor32	13
Vendedor7	13	Vendedor33	13
Vendedor8	13	Vendedor34	13
Vendedor9	13	Vendedor35	13
Vendedor10	13	Vendedor36	13
Vendedor11	13	Vendedor37	13
Vendedor12	13	Vendedor38	13
Vendedor13	13	Vendedor39	13
Vendedor14	13	Vendedor40	13
Vendedor15	12	Vendedor41	13
Vendedor16	12	Vendedor42	13
Vendedor17	13	Vendedor43	13
Vendedor18	13	Vendedor44	13
Vendedor19	13	Vendedor45	13
Vendedor20	13	Vendedor46	13
Vendedor21	13	Vendedor47	13
Vendedor22	13	Vendedor48	13
Vendedor23	13	Vendedor49	13
Vendedor24	12	Vendedor50	13
Vendedor25	13	Vendedor51	13

Tabela 5.4 – Quantidade de compradores com que cada vendedor interagiu durante as rodadas de treinamento

Comprador	Quantidade de vendedores	Fator de correção (n/50)
Comprador0	52	1
Comprador1	52	1
Comprador2	50	1
Comprador3	52	1
Comprador4	52	1
Comprador5	52	1
Comprador6	52	1
Comprador7	52	1
Comprador8	52	1
Comprador9	52	1
Comprador10	52	1
Comprador11	52	1
Comprador12	52	1

Tabela 5.5 – Quantidade de vendedores com que cada comprador interagiu durante as rodadas de treinamento

5.3.4 Conclusão dos testes comparativos

O método proposto, apesar de não ser o melhor em nenhum momento, mostrou uma estabilidade maior nos dois testes comparativos. Esta regularidade se deveu a uma queda de desempenho menos acentuada a dos outros métodos à medida que a exposição à oscilação de comportamento aumentava nos cenários testados. O nível de informação presente no ambiente também parece não influenciar no desempenho do método proposto - teve o segundo desempenho geral nos dois testes comparativos realizados.

Quanto aos demais métodos, os testes mostraram que o sucesso do método FCA2 em ambientes de oscilação de comportamento parece depender da quantidade de informação existente. Esta suposição é razoável de ser feita em função deste comportamento ser uma propriedade comum aos algoritmos de Filtragem Colaborativa, entretanto, esta propriedade pareceu não ser preservada pelo método FCA3 neste tipo de ambiente, em razão da queda acentuada de desempenho com o aumento da quantidade de informação.

Em função dos comportamentos distintos de FCA2 e FCA3 em cada teste, e de nenhum método ter sido significativamente melhor, consideramos como trabalhos futuros, a replicação destes experimentos na forma de experimentos de campo, com usuários reais, de forma a validar e refinar as conclusões obtidas inicialmente.

5.4 Outros trabalhos

Nesta seção realizamos análises comparativas adicionais, desta vez entre o método proposto e outras abordagens que não foram testadas nos experimentos da seção 5.3.

5.4.1 Sistema ReGreT

O modelo de reputação implementado no sistema ReGreT proposto por Sierra e Sabater [9, 32] é orientado para comunidades eletrônicas (ambientes de comércio eletrônico e comunidades de agentes) complexas onde as relações sociais desempenham grande importância. Trata-se de uma abordagem bem mais complexa do que o método proposto neste trabalho. O ReGreT trabalha com diferentes dimensões de reputação (individual, social e ontológica) e fontes de informação (experiências diretas, reputação por testemunhos e relacionamentos sociais), que permitem ao agente combinar os valores de reputação e

experiências diretas em uma única medida de Confiança, calculada de acordo com a quantidade de informação disponível para o agente ativo. As dimensões estabelecem a forma de cálculo de valores de reputação e de confiança, enquanto as fontes de informação estabelecem o conjunto de informações que podem ser usados para a realização destes cálculos.

Conforme visto no capítulo 3, o método proposto se inspirou em alguns dos elementos definidos no ReGreT, como as Relações de Base utilizadas na Dimensão Individual, e os Aspectos Comportamentais Complexos e Simples da dimensão Ontológica, entretanto se concentrou no cálculo de reputação por testemunho, que no ReGreT é implementado como um dos tipos de reputação da Dimensão Social. Além da Reputação por Testemunho, o ReGreT combina outros dois tipos de Reputação (Sistema e Vizinhança) em um único valor de Reputação.

Com relação à Dimensão Ontológica, a estrutura ontológica do método proposto permite apenas um Aspecto Complexo ligado a vários Aspectos Comportamentais Simples, enquanto no ReGreT é possível combinar um Aspecto Complexo a outros Complexos formando subgrafos.

Com relação ao cálculo de reputação por testemunho, existem várias diferenças entre o método proposto e o ReGreT, em razão do método proposto utilizar a técnica de FCA ao invés dos mecanismos do ReGreT, conforme descrito a seguir.

5.4.1.1 Seleção de testemunhas

O ReGreT seleciona as testemunhas através de sociogramas e heurísticas específicas. Sociogramas representam os relacionamentos sociais e a visão social que o agente tem da comunidade ou ambiente que habita. São grafos onde os nós são os agentes do sistema e os arcos são os relacionamentos existentes entre os agentes com os tipos e a intensidade associados. O ReGreT trabalha com 3 tipos de relacionamentos: cooperação – caracterizada pela troca de informações sinceras entre os agentes e a predisposição de cada um em ajudar ao outro; competição – onde os agentes perseguem os mesmos objetivos e recursos e buscam obter vantagens uns sobre os outros mentindo ou escondendo informações; comercial - caracterizada pela realização de transações comerciais, podendo ocorrer tanto com os relacionamentos de competição quanto de colaboração.

O conjunto inicial de testemunhas é formado ao se identificar no sociograma os agentes que tiveram relacionamentos do tipo comercial com o agente alvo da transação, mas este conjunto pode ser muito grande e sofrer do problema de evidência correlacionada (*correlated evidence*) – que consiste na unificação da forma de pensar pelo compartilhamento de informações ou a partir de um mesmo fato propagado por diferentes testemunhas. Nestes casos, a maior quantidade de testemunhos sobre um agente não pode ser considerada como garantia de confiança no agente. Considerando que também existe um custo para pedir testemunhos, a aplicação de heurísticas sobre o conjunto inicial permite formar grupos de agentes que mais interagiram entre si e pedir testemunhos somente àqueles mais representativos de cada grupo, evitando assim o problema de evidência correlacionada.

A heurística utilizada no ReGreT é baseada na proposta em [49]. Ela identifica os componentes do grafo, ou seja, subgrafos que sejam maximamente conectados, e para cada componente seleciona os nós com maior grau (com maior número de arcos incidentes) ou os pontos de articulação, aqueles cuja remoção resultaria na divisão do componente na maior quantidade de novos componentes.

O método proposto neste trabalho seleciona as testemunhas (avaliadores) com maior similaridade das opiniões passadas através da Correlação de Pearson [25]. Para isto usa todos os avaliadores que tiveram avaliações em comum com o usuário ativo e a sobre esta distribuição de avaliações calcula-se o Coeficiente de Pearson.

5.4.1.2 Credibilidade das testemunhas

Para garantir o nível de confiabilidade da estimativa, o ReGreT avalia a credibilidade de cada testemunha com base em dois métodos:

socialCr: Através da análise dos relacionamentos sociais entre a partes envolvidas (testemunha, agente avaliado e agente ativo) e da aplicação de regras nebulosas [50] para conferir um grau de credibilidade. As regras nebulosas assumem a forma de estruturas condicionais que avaliam os tipos de relacionamento entre testemunha e agente avaliado, bem como a intensidade desta relação e, a partir daí atribuem o grau de confiabilidade ao testemunho emitido, conforme mostrado no exemplo a seguir:

se $coop(w,b) = \text{alto}$ **então**

$$socialCr(a, w, b) = \text{muito_baixo}$$

onde w é a testemunha, b o agente avaliado, e a é o agente solicitante do testemunho.

A regra acima indica que, se a relação entre a testemunha e o agente avaliado for de alta cooperação, sob a perspectiva do solicitante a credibilidade do testemunho é muito baixa.

Não faz parte do escopo do ReGreT criar e manter as informações dos sociogramas, mas um método para a identificação dinâmica destes relacionamentos foi proposto [51]. Os autores alegam que modelos de reputação como o ReGreT, que exploram a influência dos relacionamentos sociais no cálculo de confiança, podem ser beneficiados com a abordagem proposta, que consiste em mapear um modelo de identificação de relacionamentos genérico (baseado no Framework SMART [52]) a uma ontologia que descreve as características de um mercado eletrônico (ABMM) e interpretar os valores de confiança com base nos relacionamentos descobertos neste processo.

infoCr: Avaliação da precisão dos testemunhos anteriores feitos pela testemunha atual. Os valores dos testemunhos feitos anteriormente são comparados com os valores das experiências diretas entre o agente ativo e o agente recomendado pela testemunha sendo avaliada. Quanto maior for a diferença, menor será a credibilidade da testemunha.

Para garantir o nível de confiabilidade da estimativa de reputação, o método proposto reduz a chance de serem selecionados falsos bons vizinhos dentro do conjunto de testemunhas através da realização de ajustes sobre o histórico de avaliações entre as testemunhas e o usuário ativo, com base na similaridade das preferências de reputação deles. Somente após este procedimento, calcula os coeficientes de similaridades entre o usuário ativo e as testemunhas, e posteriormente a estimativa de reputação.

5.4.1.3 Cálculo de agregação de reputação

O ReGreT usa a credibilidade de cada testemunha como peso de avaliação de reputação enquanto o método proposto utiliza como peso a similaridade calculada através da Correlação de Pearson.

Como último comentário comparativo entre o ReGreT e o método proposto neste trabalho, destacamos que, se por um lado, os processos de seleção e verificação de credibilidade de testemunhos

são mais otimizados no ReGreT, isto irá requerer um conhecimento significativamente maior do ambiente em que habitam [11]. Os agentes que necessitarem usar efetivamente o tipo de reputação por testemunho terão o ônus da manutenção de sociogramas que contenham, no mínimo, as relações sociais dos agentes que irão interagir, tarefa que pode se tornar muito onerosa em ambientes com grande quantidade de agentes e volume de transações.

Por outro lado, o método proposto, se aplicado neste domínio, não oneraria os agentes solicitantes com a manutenção destas informações, uma vez que o método de seleção de testemunhas se baseia exclusivamente em informações já presentes nos próprios testemunhos (valores de reputação e as preferências de reputação), não requerendo a manutenção de nenhuma estrutura de dados adicional.

5.5 Modelo de credibilidade de testemunhas

O trabalho de Huynh et al. [11] propõe um método genérico para detecção de imprecisões nos testemunhos de reputação e atualização da credibilidade das testemunhas de acordo com o grau de imprecisão observado. As imprecisões podem ocorrer em razão dos diferentes pontos de vista que cada testemunha tem das interações ou do comportamento tipicamente egoísta dos agentes observado em sistemas multiagentes abertos, que faz com que os agentes escondam ou forneçam falsas informações para atender aos seus próprios interesses. A imprecisão é definida através da magnitude das diferenças apuradas entre o testemunho fornecido e o desempenho real do agente avaliado.

Um modelo de credibilidade de testemunhas é proposto que represente estas imprecisões com base nas próprias experiências do agente solicitante, sem requerer qualquer conhecimento adicional do domínio (ex.: as relações sociais entre os agentes). Estas imprecisões são usadas como medidas de peso de relevância sobre os testemunhos durante o processo de geração do valor de reputação. Apesar de embasarem a aplicação do modelo de credibilidade em um modelo específico de confiança denominado FIRE, os autores afirmam que ele pode ser combinado com qualquer modelo de confiança.

Em linhas gerais o modelo FIRE [11] trabalha com dois componentes básicos: Confiança de Interações que calcula a confiança a partir das experiências diretas entre os agentes; Reputação por Testemunho, que consiste na obtenção de testemunhos de reputação sobre o agente alvo da interação. O

componente de interações diretas calcula a média das avaliações passadas do agente solicitante sobre o agente alvo, ponderada por fatores que ajustam os valores de avaliação com base no quão recente é a avaliação feita, evitando que experiências antigas (vencidas) reflitam de forma inadequada o comportamento atual do agente avaliado. Outra característica deste componente é a atribuição de valores de reputação por itens de contrato, como por exemplo: preço, qualidade, entrega, etc.

O componente de reputação permite que os agentes conhecidos pelo agente solicitante indiquem outros agentes capazes de responder à solicitação de testemunho, ou eles próprios fornecem os testemunhos sobre o agente alvo. O cálculo da reputação é obtido através da média das avaliações dos testemunhos, ponderada com base no fator de credibilidade apurado pelo modelo proposto.

Se aplicado no domínio de sistemas multiagentes, o método proposto neste trabalho reconheceria como agentes conhecidos aqueles que avaliaram usuários em comum ao ativo.

A forma de cálculo da credibilidade é a mesma utilizada no componente de interações diretas, só que com base no histórico de avaliações oferecidas pelas testemunhas sobre outros agentes. O modelo de credibilidade proposto preconiza que o testemunho seja um serviço oferecido por cada agente do sistema, ou seja, independente do modelo de confiança do agente ele deve ser capaz de usá-lo para fornecer testemunhos a qualquer agente solicitante sobre a confiabilidade de outros agentes. O agente solicitante pode então comparar o desempenho efetivo do agente alvo com o valor fornecido pelo serviço de testemunho de cada testemunha. Quanto maior a diferença menor será a credibilidade da testemunha, e menor será o peso futuramente utilizado no cálculo de reputação do componente de Reputação por Testemunho. Os autores definem a fórmula abaixo para o cálculo da credibilidade de cada testemunha e agente alvo:

$$v_w = \begin{cases} 1 - |v_k - v_a| & \text{se } |v_k - v_a| < \tau \\ -1 & \text{se } |v_k - v_a| \geq \tau \end{cases}$$

onde v_w é o valor de credibilidade de um agente alvo b avaliado por uma testemunha w , na faixa de -1 a 1, v_k e v_a são as avaliações reais e testemunhadas da confiabilidade do agente alvo definidos na faixa de -1 a 1, e τ é um limite de tolerância de imprecisão definido para o sistema podendo assumir um valor

entre 0 e 2 (diferença máxima possível dado que os valores de v_k e v_a estão entre -1 e 1). Caso o agente solicitante não tenha tido nenhuma experiência passada com uma determinada testemunha, será atribuído um valor padrão de credibilidade, que irá refletir a política definida no ambiente para uso de testemunhos de testemunhas que nunca encontrou antes.

Da mesma forma que neste método, o método proposto neste trabalho também faz algo equivalente a uma verificação de credibilidade sobre o histórico de avaliações em comum entre cada testemunha e o usuário ativo. Só que a verificação da credibilidade é feita com base na similaridade das preferências de reputação da testemunha e do solicitante na época, e não na diferença de desempenho real e testemunhado sobre o elemento avaliado.

Uma das limitações atuais deste trabalho é que o limite de intolerância (t) não é dinamicamente aferido de acordo com a comunidade e os agentes presentes nela. Além disto, precisa ser definido cuidadosamente para não classificar de forma incorreta as imprecisões, porque as diferenças com valor acima deste limite implicarão na classificação do testemunho como impreciso e o agente que o forneceu como mentiroso, sendo penalizado com o valor de credibilidade mínimo de -1. Portanto, um valor muito baixo pode classificar incorretamente como mentirosos agentes honestos, e um valor muito alto pode classificar incorretamente como honestos, agentes mentirosos.

Capítulo 6 Conclusões

Neste trabalho é proposto um método para que o cálculo de estimativas de reputação personalizadas baseadas em FCA funcionem de forma adequada na presença de comportamento variável das entidades avaliadas.

Foram realizados testes de recomendação em cenários de grande oscilação de comportamento por parte dos vendedores, onde o método proposto se comportou adequadamente. Em todos os cenários o método foi capaz de priorizar recomendações de vendedores tipo Good, possibilitando a apuração de lucro aos compradores enquanto houvesse vendedores do tipo Good disponíveis para transacionar, pouco sofrendo com a infiltração de vendedores Bad entre as recomendações iniciais.

Foram estudadas abordagens correlatas que tinham como preocupação em comum conferir um grau de confiabilidade adequado para o conjunto de vizinhos selecionados, e a apuração dos valores de reputação mais confiáveis. Dois experimentos comparativos entre o método proposto e estas abordagens foram realizados, cada qual diferenciando-se na quantidade de avaliações presentes no ambiente. O método proposto apresentou o segundo melhor desempenho geral, além de ter o desempenho mais estável entre as abordagens testadas nos dois experimentos. Esta regularidade se deveu a uma queda de desempenho menos acentuada aos demais à medida que a exposição à oscilação de comportamento aumentava. Outra observação importante de ser feita é que o nível de esparsidade presente no ambiente também não influenciou no desempenho do método proposto, uma vez que seu desempenho se manteve sempre estável nos dois experimentos.

6.1 Contribuições

A contribuição central deste trabalho foi a definição de um método para geração de estimativas de reputação personalizadas que funcione de forma confiável em ambientes onde as entidades avaliadas possam mudar de comportamento entre diferentes avaliações.

Mostramos nos experimentos que, com a inclusão de preferências de reputação e aplicação de transformações sobre a matriz de avaliações, o método proposto pode gerar recomendações confiáveis em

ambientes de alta oscilação de comportamento em cenários com variada quantidade de avaliações de reputação registradas. Apesar de nos experimentos o método proposto não ter apresentado desempenho significativamente melhor em relação aos métodos correlatos testados, verificamos que manteve um desempenho mais estável em relação aos demais em ambientes com variados graus de esparsidade.

O método proposto pode ser complementado por Modelos de Confiança, e ser empregado em qualquer sistema onde haja necessidade de fornecimento de serviços, produtos ou conteúdo com confiabilidade (mercados eletrônicos, redes de compartilhamento ponto a ponto), envolvendo interações entre agentes humanos ou computacionais.

Apesar de não ser o objetivo inicial deste trabalho, foi observado que algumas características do método proposto (transformações das matrizes de avaliação e encapsulamento das preferências de reputação), por diminuírem o nível de informação existente no ambiente, podem criar as condições necessárias para melhor proteger o sistema contra manipulações por grupos de clientes e fornecedores desonestos.

Uma outra contribuição deste trabalho foi a revisão bibliográfica de sistemas de recomendação e de reputação, descrevendo seus benefícios e limitações, principais áreas de estudo passado e futuro, e como se complementam.

6.2 Limitações

Por ser baseado nas técnicas de filtragem colaborativa, o método proposto herda alguns dos problemas inerentes destas técnicas, como o problema da primeira avaliação, a esparsidade de informações de avaliação, e o tempo de geração de recomendações.

A questão de desempenho (tempo de geração de recomendações) se acentua no método proposto, tendo em vista que a matriz de avaliações proposta mantém além dos valores de reputação as preferências de reputações, e que para cada geração de lista de recomendações é criada uma matriz de avaliação temporária transformada de acordo com estas referências de reputação. Uma das alternativas de otimização do método proposto é possibilidade de manutenção de matrizes transformadas não transientes

para cada usuário, atualizadas pontualmente a cada nova avaliação inserida na matriz de avaliação central que envolva algum dos vendedores ou avaliadores existentes na matriz transformada.

O método atualmente implementado não permite o registro de múltiplas avaliações entre um mesmo cliente e fornecedor, mantendo apenas a última avaliação na matriz de avaliações. Quando o fizer, precisarão ser definidos critérios de seleção de avaliações (dentro das janelas de tempo) que também levem em consideração as preferências de reputação associadas a cada uma delas. Médias e outras funções de agregação poderão ser aplicadas, desde que tanto sobre os valores de avaliação, quanto sobre os quesitos de avaliação, que também podem variar com o tempo. Também deve ser prevista a evolução das preferências de reputação, com a criação de regras para tratar a criação ou exclusão de quesitos de avaliação ao longo do tempo nos cálculos de similaridade entre as preferências de reputação dos usuários ativos e vizinhos.

Devido à natureza centralizada das técnicas de filtragem colaborativa, o método proposto, ao ser aplicado em sistemas multiagentes ou redes ponto-a-ponto totalmente descentralizadas, precisa contar com alguns agentes ou nós centralizadores para concentrar as avaliações e o fornecimento das recomendações. Isto não impede que outros modelos distribuídos de reputação ou de confiança sejam implementados e convivam com o método proposto.

6.3 Trabalhos futuros

A implantação do método proposto em sistemas multiagentes abertos onde convivam diferentes abordagens de reputação é um importante objeto de estudo futuro. A criação de agentes especializados no fornecimento de serviços de estimativa de reputação, a criação de mecanismos de incentivos genéricos para que estes agentes recebam testemunhos de outros agentes, e a necessidade de garantir a imparcialidade destes agentes, dado o comportamento egoísta (*self-interested*) usualmente observado nestes ambientes [11], são algumas das ações necessárias para viabilizar a aplicação do método proposto nestes ambientes.

O mecanismo básico de Filtragem de Agrupamentos, voltado para as situações de avaliação desonesta positiva (*Ballot Stuffing*) e discriminação positiva de compradores, foi testado e obteve melhor desempenho em relação ao método proposto, apesar disto, devido aos experimentos não terem simulado as condições originais de teste deste mecanismo, não foi possível concluir se o desempenho deste mecanismo é melhor do que o método proposto nem realizar testes com o mecanismo avançado de Filtragem de Agrupamentos voltado para cenários de inundação de avaliações desonestas. Com a elaboração de novos experimentos onde os avaliadores, e não mais os fornecedores (avaliados), sejam as variáveis controladas pelo experimento, será possível comparar efetivamente o desempenho do método proposto não só com o mecanismo básico mas também com o mecanismo estendido.

Os novos experimentos acima indicados permitirão verificar se, na prática, se as propriedades de encapsulamento e transformação de informações de reputação realmente conferem maior proteção ao sistema.

Em função de nos testes comparativos as diferenças de desempenho entre os métodos não terem sido estatisticamente significantes, da necessidade de confirmar a estabilidade do método proposto em ambientes com variados graus de esparsidade, e para melhor compreender por que o aumento de informações de avaliação influenciou de forma distinta os desempenhos dos métodos de Filtragem de Agrupamentos e Peso de Relevância, consideramos necessário realizar estas investigações também no contexto dos novos experimentos citados acima, para que as conclusões obtidas inicialmente sejam validadas.

Referências Bibliográficas

- [1] Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P. and Riedl, J.: GroupLens: An Open Architecture for Collaborative Filtering of Netnews. In *Proceedings of 1994 Conference on Computer Supported Cooperative Work*, (Chapel Hill, North Carolina, 1994). ACM, 175-186.
- [2] Amazon. <http://www.amazon.com>.
- [3] iTunes. (June 05, 2006), <http://www.apple.com/itunes/>.
- [4] Dellarocas, C.: Immunizing Online Reputation Reporting Systems Against Unfair Ratings and Discriminatory Behavior. In *Proceedings of the Conference on Electronic Commerce*, (Minneapolis, Minnesota, 2000). ACM, 150-157.
- [5] Herlocker, J.L., Konstan, J.A., Borchers, A. and Riedl, J.: An algorithm framework for performing collaborative filtering. In *Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, (1999).
- [6] Dellarocas, C.: Analyzing the Economic Efficiency of eBay-like Online Reputation Reporting Mechanisms. In *3rd ACM Conference on Electronic Commerce*, (Tampa, FL, 2001).
- [7] Paul Resnick, R.Z.: Trust Among Strangers in Internet Transactions: Empirical Analysis of eBay's Reputation System. in Baye, M. ed. *The Economics of the Internet and ECommerce*. JAI Press, 2002.
- [8] Resnick, P., Zeckhauser, R., Friedman, E. and Kuwabara, K.: Reputation Systems. *Communications of the ACM*, 43, 12 (2000), 45-48.
- [9] Jordi Sabater, C.S.: Reputation and Social Network Analysis in Multi-Agent Systems. In *AAMAS'02*, (Bologna, Italy., 2002).
- [10] Lik Mui, A.H., Mojdeh Mohtashemi: Notions of Reputation in Multi-Agents Systems: A Review. In *AAMAS'02*, (Bologna, Italy., 2002).
- [11] Huynh, T.D., Jennings, N.R. and Shadbolt, N.R.: On Handling Inaccurate Witness Reports. In *Proceedings of the 8th International Workshop on Trust in Agent Societies*, (Utrecht, The Netherlands, 2005), 63-77.
- [12] John O'Donovan, B.S.: Trust in Recommender Systems. In *IUI'05*, (San Diego, California, 2005).
- [13] Zacharia, G., Moukas, A. and Maes, P.: Collaborative Reputation Mechanisms in Electronic Marketplaces. In *Proceedings of the HICSS-99 Conference, ElectronicCommerce Minitrack*, (Hawaii, 1999).
- [14] Torres, R.: *Personalização na Internet*. (Novatec Editora, São Paulo, 2004).
- [15] Resnick, P. and Varian, H.R.: Recommender Systems. *COMMUNICATIONS OF THE ACM*, 40, 3 (1997), 56-58.
- [16] Konstan, J.A.: Introduction to recommender systems: Algorithms and Evaluation. *ACM Trans. Inf. Syst.*, 22, 1 (2004), 1-4.

- [17] Goldberg, D., Nichols, D., Oki, B.M. and Terry, D.: Using collaborative filtering to weave an information tapestry. *COMMUNICATIONS OF THE ACM*, 35, 12 (1992), 61-70.
- [18] Shardanand, U. and Maes, P.: Social information filtering: algorithms for automating “word of mouth”. In *SIGCHI conference on Human factors in computing systems*, (Denver, Colorado, United States, 1995). ACM Press/Addison-Wesley Publishing Co., 210-217.
- [19] Schafer, J.B., Konstan, J. and Riedi, J.: Recommender Systems in E-Commerce. In *Proceedings of the 1st ACM conference on Electronic commerce*, (Denver, Colorado, United States, 1999). ACM Press, 158-166.
- [20] Amazon. (January 02, 2006), <http://amazon.com>.
- [21] eBay. (January 02, 2006), <http://www.ebay.com>.
- [22] Schafer, J.B.: MetaLens: A Framework for Multi-Source Recommendations. in *Computer Science*. University of Minnesota, Minneapolis, 2001.
- [23] Lam, S.T.K. and Riedl, J.: Shilling Recommender Systems for Fun and Profit. In *Proceedings of the 13th international conference on World Wide Web*, (New York, NY, 2004). ACM Press, 393-402.
- [24] John S. Breese, D.H., Carl Kadie: Empirical Analysis of Predictive Algorithms for Collaborative Filtering. Microsoft Corporation (1998).
- [25] Barbetta, P.A.: O coeficiente de correlação linear de Pearson. in *Estatística Aplicada às Ciências Sociais*, Florianópolis, 2002, 275-282.
- [26] Jøsang, A., Ismail, R. and Boyd, C.: A Survey of Trust and Reputation Systems for Online Service Provision. *Decision Support Systems* (2006).
- [27] McKnight, D.H. and Chervany, N.L.: The Meanings of Trust *Technical Report MISRC Working Paper Series 96-04*. University of Minnesota, Management Information Systems Research Center (1996).
- [28] Sabater, J. and Sierra, C.: Reputation and Social Network Analysis in Multi-Agent Systems. In *Proceedings of the first international joint conference on Autonomous agents and multiagent*, (Bologna, Italy, 2002). ACM Press, 475-482.
- [29] epinions. (June 13, 2006), <http://www.epinions.com/>.
- [30] Dingedine, R., Freedman, M.J. and Molnar, D.: Accountability Measures for Peer-to-Peer Systems. in *Peer-to-Peer: Harnessing the Power of Disruptive Technologies*. O'Reilly Publishers, 2000.
- [31] Jordi Sabater, C.S.: Review on Computational Trust and Reputation Models. *Artificial Intelligence Review*, 24, 1 (2005), 33-60.
- [32] Sabater, J.: Trust and reputation for agent societies *Institut d'Investigació en Intel·ligència Artificial*, Bellaterra, Catalonia, Spain. (2003).

- [33] Mudhakar Srivatsa, L.X., Ling Liu: TrustGuard: Countering Vulnerabilities in Reputation Management for Decentralized Overlay Networks. In *14th international conference on World Wide Web*, (Chiba, Japan, 2005).
- [34] Chrysanthos Dellarocas, P.R.: Online Reputation Mechanisms A Roadmap for Future Research. In *First Interdisciplinary Symposium on Online Reputation Mechanisms*, (Cambridge, MA, 2003), 1-16.
- [35] H. Chi Wong, K.S.: A Taxonomy of Middle-agents for the Internet. In *International Conference on Autonomous Agents and Multi-Agent Systems (ICMAS-00)*, (2000).
- [36] ebay. <http://www.ebay.com>.
- [37] The Internet Movie Database. <http://www.imdb.com>.
- [38] MovieLens. (January 1 2006, <http://www.movielens.org>).
- [39] Anthony Chavez, P.M.: Kasbah: An Agent Marketplace for Buying and Selling Goods. In *First International Conference on the Practical Application of Intelligent Agents and Multi-Agent Technology*, (London, UK, 1996). Practical Application Company, 75-90.
- [40] Androutsellis-Theotokis, S.: A Survey of Peer-to-Peer Content Distribution Technologies. *ACM Comput. Surv.*, 36, 4 (2004), 335-371.
- [41] Dellarocas, C.: Contractual Agent Societies: Negotiated shared context and social control in open multi-agent systems. In *4th International Conference on Multi-Agent Systems*, (2000).
- [42] Zoran Milosevic, A.J., Theo Dimitrakos, Mary Anne Patton: Discretionary Enforcement of Electronic Contracts. In *EDOC '02*, (2002), 39-50.
- [43] Michal Morciniec, M.S., Brian Monahan: *Towards Regulating Electronic Communities with Contracts*. Montreal, 2001).
- [44] Owen, S. Taste. 0.7. <http://sourceforge.net/projects/taste/>.
- [45] Sabater, J.: Evaluating the ReGreT system. *Applied Artificial Intelligence*, 18, 9/10 (2004), 797-813.
- [46] Lopes, A. and Garcia, A.C.B.: Applying Collaborative Filtering to Reputation Domain: a model for generating more precise reputation estimates in case of changing behavior by the rated agents. In *AAMAS Workshop on Trust in Agent Societies (9th)*, (Hakodate, Japan, 2006). To be published.
- [47] Barbeta, P.A.: Testes de comparação entre duas amostras. in *Estatística Aplicada às Ciências Sociais*. Editora da UFSC, Florianópolis, 2002, 227.
- [48] P. Macnaughton-Smith, W.T.W., M.B.Dale, L.G. Mockett: Dissimilarity analysis: A New Technique of Hierarchical Sub-division. *Nature*, 202 (1964), 1034-1035.
- [49] Per Hage, F.H.: *Structural Models in Anthropology*. (Cambridge University Press, 1983).
- [50] Zadeh, L.A.: Fuzzy logic and approximate reasoning. *Synthese*, 30 (1975), 407-428.

- [51] Ashri, R., Ramchurn, S.D., Sabater, J., Luck, M. and Jennings, N.R.: Trust Evaluation Through Relationship Analysis. In *Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems*, (The Netherlands, 2005). ACM Press, 1005-1011.
- [52] Michael Luck, M.d.I.: *Understanding Agent Systems*. (Springer, 2004).

Anexo A Modelagem do método proposto

Diagrama de Classes

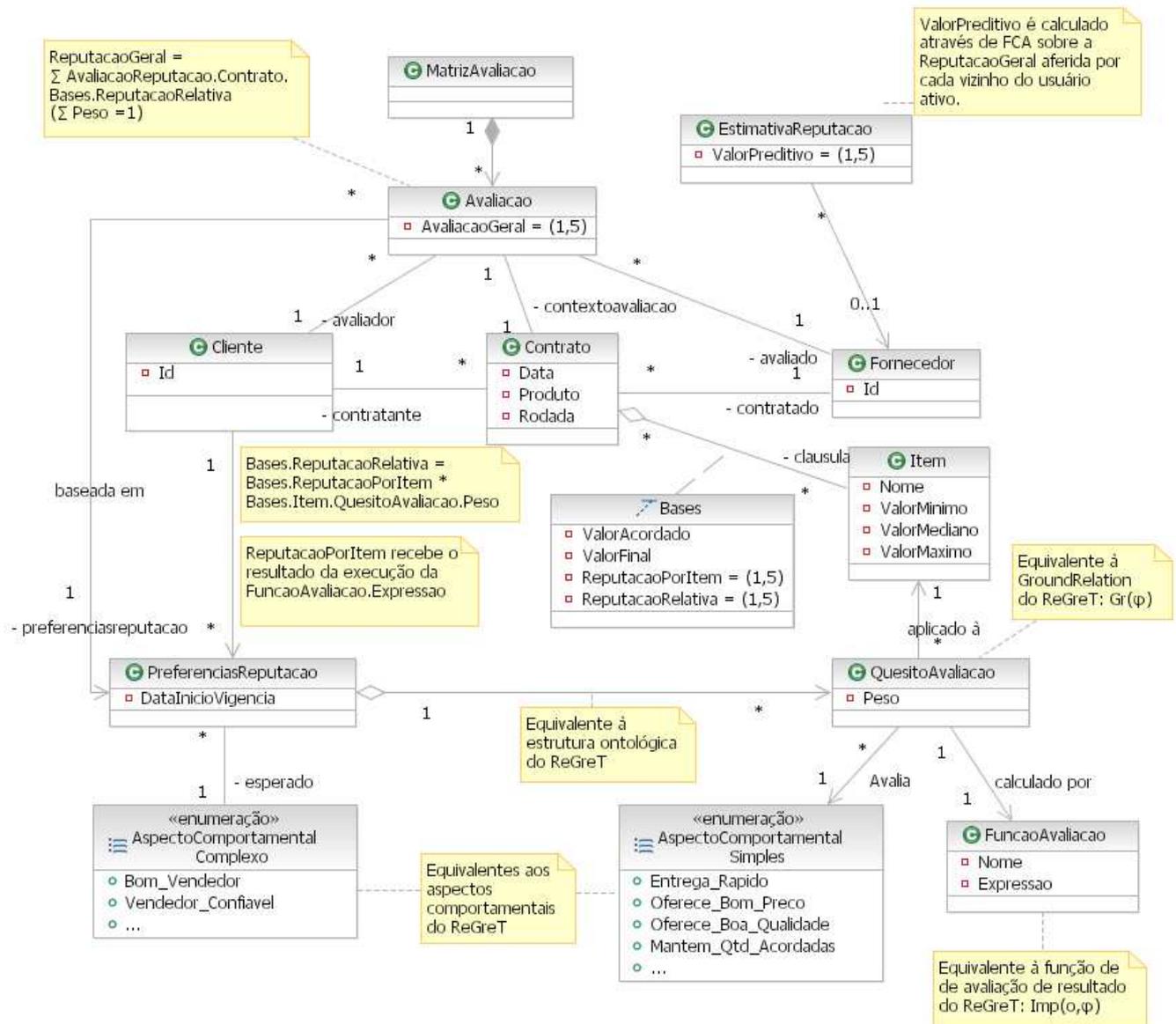


Figura A.1 - Modelo conceitual completo (notação UML 2.0)

Para efeito de clareza das explicações, a descrição deste diagrama, abaixo apresentada, também foi inserida no capítulo “Método Proposto”.

Os clientes e fornecedores são entidades devidamente registradas no ambiente que realizam transações com objetivo de atendimento de necessidades financeiras e de consumo. Ao final de transação o cliente emite uma avaliação do fornecedor. Cada avaliação está associada a um contrato e à preferência de reputação vigente naquele momento. Os contratos são estabelecidos entre clientes e fornecedores para

registrar as bases acordadas (valores) para realização das transações. Cada contrato é constituído por cláusulas associadas aos itens acordados, como preço, prazo de entrega, quantidade e qualidade.

As preferências de reputação mudam ao longo tempo e são empregadas pelo cliente na avaliação dos fornecedores contratados. Elas são compostas por quesitos de avaliação. Cada item de contrato está associado a um quesito de avaliação que possui um peso e forma de avaliação. Os quesitos de avaliação indicam como os desvios dos valores finais em relação aos valores acordados nas bases do contrato influenciam negativamente ou positivamente em cada aspecto comportamental. Neste sentido têm função similar às Relações de Base (*Ground Relation*) definidas no sistema ReGreT [32], sendo a influência apurada através de uma função de avaliação dependente do domínio e descrita no atributo expressão da classe FuncaoAvaliacao.

O método proposto compartilha a premissa do ReGreT de que a reputação é um conceito complexo (Aspectos Comportamentais Complexos) aferido através da combinação de várias dimensões de avaliação mais simples (Aspectos Comportamentais Simples). Desta forma cada preferência de reputação é composta por vários quesitos de avaliação, que combinados com seus pesos determinam o valor de reputação geral, formando um relacionamento similar às Estruturas Ontológicas definidas no sistema ReGreT [32].

Diagrama de Atividades

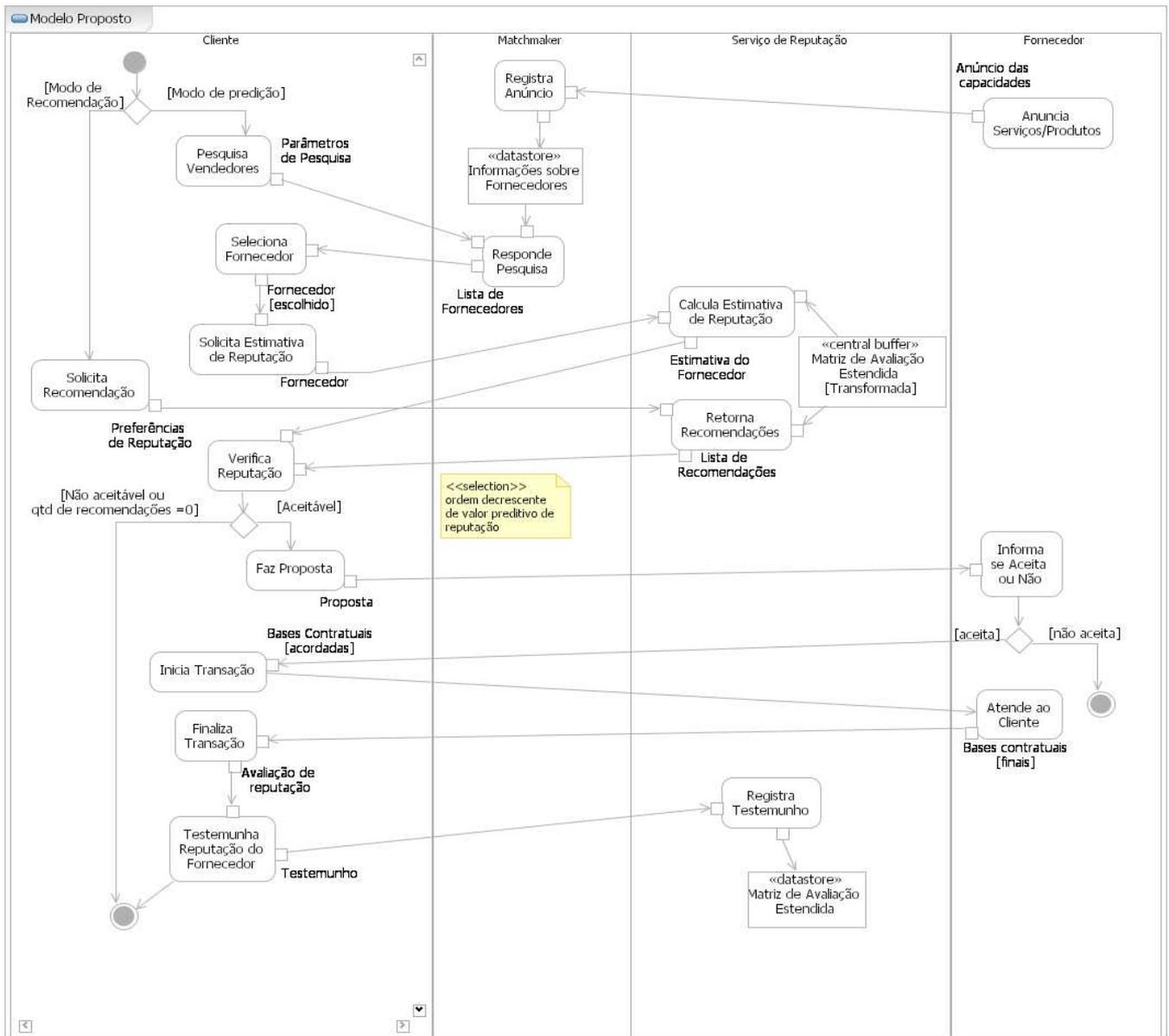


Figura A.2 - Fluxo de dados e de controle do método proposto (notação UML 2.0)

Anexo B Modelagem da simulação

Diagrama de Classes (Modelo de Análise)

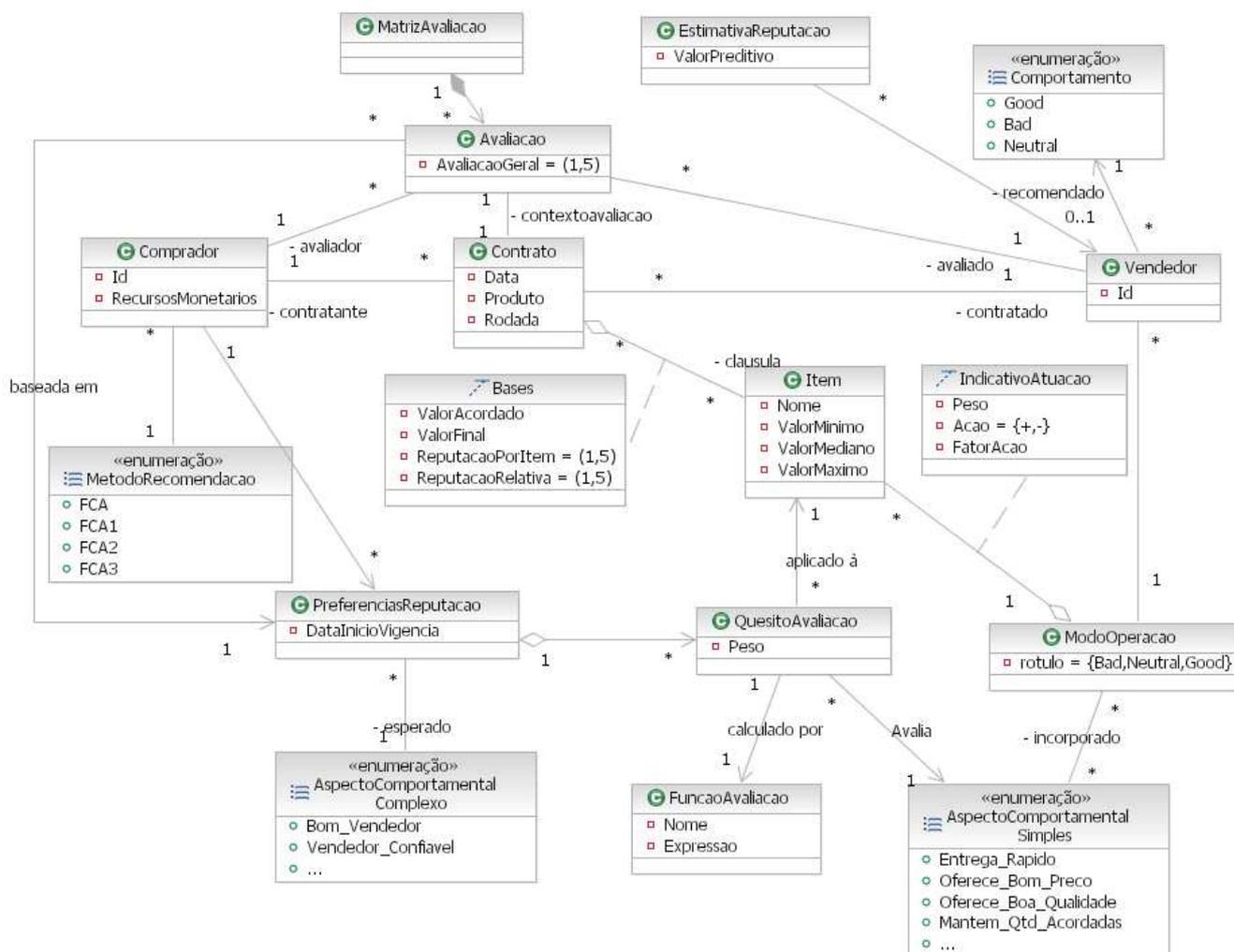


Figura B.1 - Visão conceitual da simulação (notação UML 2.0)

A simulação é representada através de um diagrama de classes que é um refinamento do modelo conceitual do método proposto descrito no capítulo 3. Alguns elementos deste modelo foram adaptados e outros simplificados para representar aspectos específicos da simulação, como as classes Comprador e Vendedor que desempenham os papéis de Cliente e Fornecedor, e a classe Contrato, que possui apenas duas cláusulas (preço e qualidade) e tem a indicação da rodada da simulação em que foi estabelecido. Além disto, são considerados alguns elementos adicionais como a classe MetodoRecomendacao, que lista o conjunto de métodos testados na simulação, e a classe Comportamento, que implementa os tipos de comportamentos que um Vendedor pode assumir na simulação.

Diagrama de Atividades

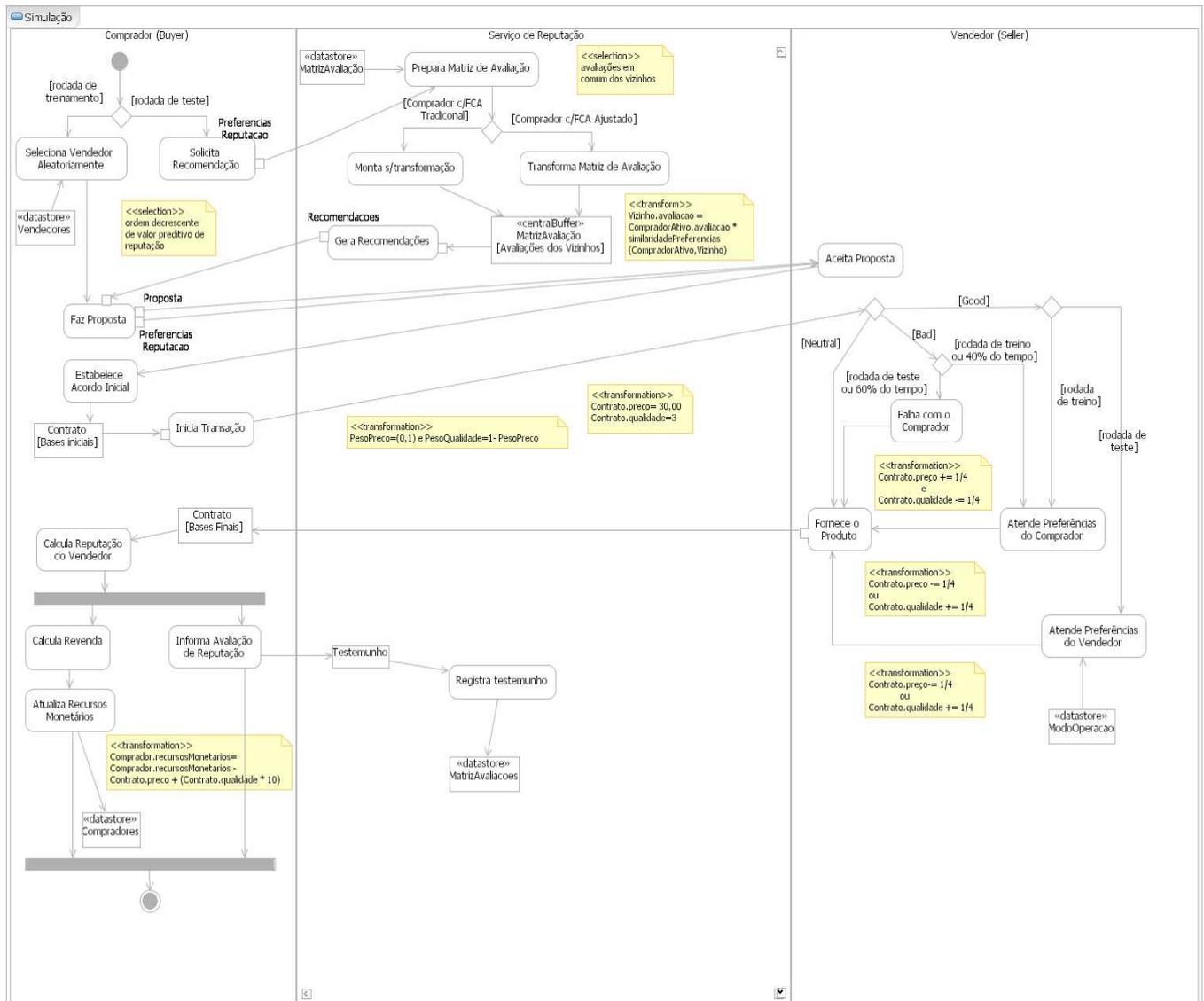


Figura B.2 - Fluxo de controle e de dados da simulação, com o Serviço de Reputação implementando o método proposto (notação UML 2.0)

Em função do sistema modelado ser uma simulação, as entidades externas que interagem com o sistema (compradores e vendedores) são simulados através de *threads* de processo do sistema, o que impede sua modelagem como atores, e a derivação de casos de uso. Desta forma, optamos por utilizar diagramas de atividades, ao invés de diagramas de caso de uso, onde as Partições de Atividade (*swimlanes*) representem não só o mecanismo de reputação, como também cada uma das entidades simuladas, e os estados de ação (*action states*) representem tarefas que estas entidades executam com o sistema.

O Serviço de Reputação é a implementação dos módulos Estimador de Reputação, Transformador, Agregador e Coletor descritos no modelo de alto nível do capítulo 3. Todas as ações de ajuste da matriz de avaliação estendida, de geração de estimativas de reputação, e de coleta de testemunhos são responsabilidade do serviço de reputação.

As funções do módulo Mediador e Publicador são implementadas pelo próprio processo central de controle da simulação (*thread main*), uma vez que os recursos de publicação e de mediação são simplificados na simulação.

A entidade simulada Comprador é responsável por solicitar estimativas de reputação, iniciar a mediação e transação com o vendedor, e por último, avaliar o vendedor com base em seu desempenho na transação. A entidade simulada Vendedor, por sua vez se responsabiliza por decidir a forma de atuação com os compradores em cada transação, por definir se mantém as bases acordadas nos contratos, e por entregar os produtos solicitados.

Diagrama de Classes (Modelo de Design)

Pacotes do Sistema



Figura B.3 – Pacotes da simulação (notação UML 2.0)

Pacote simulator.controller

Contém a classe responsável por controlar toda a simulação: iniciação, execução das rodadas de treinamento e de teste, e finalização.

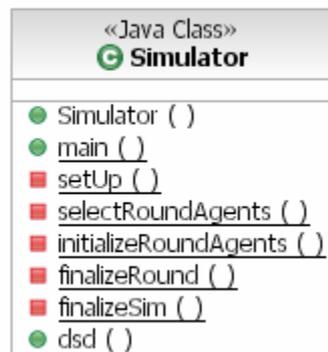


Figura B.4 – Classe do pacote simulator.controller (notação UML 2.0)

A classe Simulator é responsável pelo controle de execução geral da simulação. Se encarrega por iniciar o banco de dados onde são registradas as avaliações, por solicitar a ativação de todos os vendedores e compradores, por definir aleatoriamente o comportamento (Bad, Good ou Neutral) dos vendedores, definir o método de recomendação usado pelos compradores, aleatoriamente definir as preferências de reputação de cada comprador, e por rodada selecionar os compradores a participarem de transações. Por último se encarrega de finalizar a simulação solicitando a interrupção dos threads de vendedores e compradores e fechando o arquivo de log da simulação.

Pacote `simulator.agents`

Contém as classes responsáveis por simular os compradores e vendedores, bem como classes responsáveis pelo seu gerenciamento (localização, obtenção, ativação e interrupção).

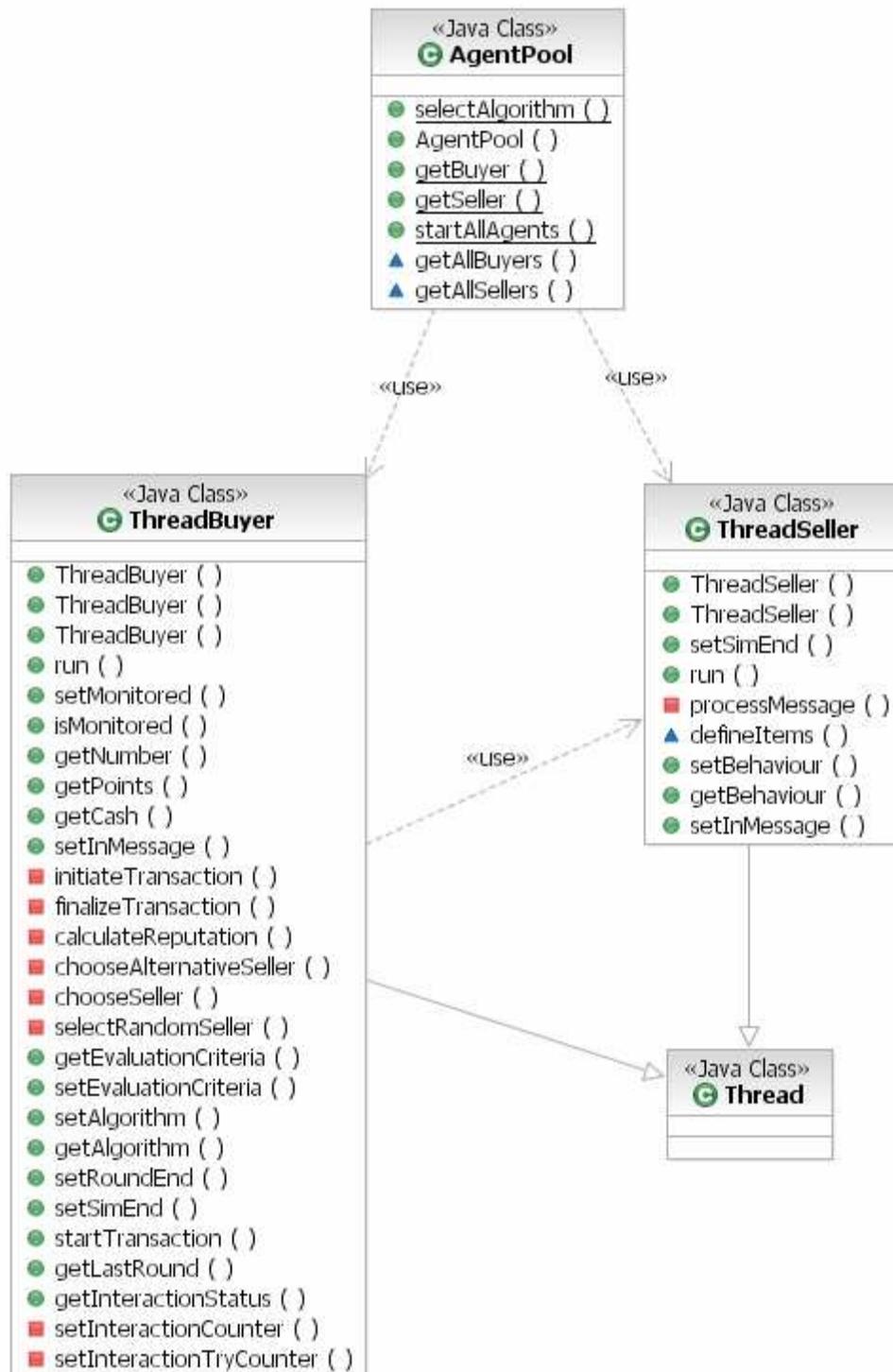


Figura B.5 – Classes do pacote `simulator.agents` (notação UML 2.0)

As classes *ThreadBuyer* e *ThreadSeller* simulam os compradores e vendedores respectivamente. São especializações da classe *Thread* do java, e por esta razão funcionam de forma ativa a partir do momento em que são iniciadas. A comunicação entre estas classes é feita de forma assíncrona através do envio, enfileiramento e processamento de mensagens do protocolo de comunicação definido para a simulação. *ThreadBuyer* é responsável por calcular a reputação de cada *ThreadSeller* que interage e por manter seus valores financeiros atualizados à medida que realiza transações e apura lucro ou prejuízo. Quando o ciclo de transação termina, atualiza os contadores de transação do sistema, e suspende sua execução até que seja acionada novamente pela classe *Simulator* na rodada seguinte. *ThreadSeller* processa a proposta de transação de acordo com seu comportamento definido, informando a contra-proposta relativa aos itens preço e qualidade do contrato estabelecido com o comprador. Ao serem informadas do término da simulação, as classes *ThreadBuyer* e *ThreadSeller* interrompem sua execução.

AgentPool é responsável por gerenciar os agentes compradores e vendedores, ativando os *threads* de execução deles, localizando-os e retornando referências de acesso para o controle de simulação solicitar a execução de transações.

Pacote simulator.control

Contém as classes que sincronizam a execução de todos os vendedores e compradores, e responsáveis pelo incremento das rodadas à medida que todas as transações sejam finalizadas.

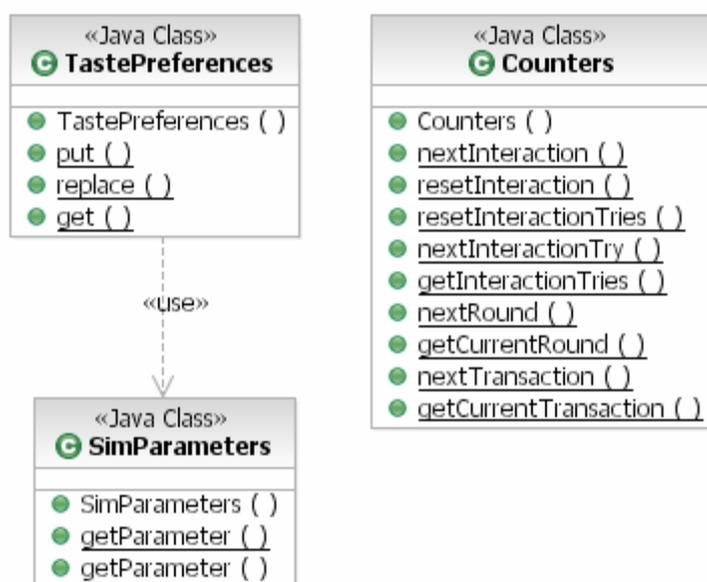


Figura B.6 – Classes do pacote simulator.control (notação UML 2.0)

A classe *TastePreferences* implementa a matriz de avaliações estendida utilizada pelo método proposto neste trabalho, mantendo as informações das preferências de reputação utilizadas pelos compradores nas avaliações dos vendedores, e mantendo ponteiros para as avaliações que são armazenadas em um banco de dados mysql versão 4.1.

A classe *SimParameters* encapsula o acesso aos parâmetros da simulação, que são mantidos em um arquivo xml cuja definição é apresentada posteriormente neste anexo.

A classe *Counters* implementa todo o controle de sincronismo da simulação, implementando contadores de transações iniciadas e terminadas, e contadores de rodadas, o que permite a classe *Simulator* identificar em que ponto se encontra a simulação, quando é o momento de iniciar uma próxima rodada, e quando a simulação deve ser encerrada.

Pacote *simulator.util*

Contém as classes utilitárias do sistema, que fornecem serviços básicos utilizados pelas classes dos demais pacotes.

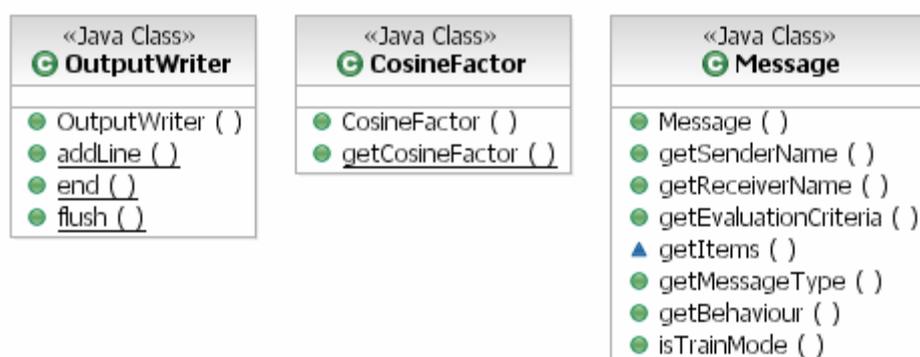


Figura B.7 – Classes do pacote *simulator.util* (notação UML 2.0)

A classe *OutputWriter* implementa o armazenamento das informações da simulação em arquivos, registrando informações sobre as transações realizadas, como por exemplo: os valores acordados e realizados, as partes envolvidas, a rodada em que a transação se realizou e qual o método de recomendação utilizado em cada transação.

A classe *CosineFactor* implementa a função de co-seno utilizada pelo método proposto para identificar a similaridade entre as preferências de reputação do vizinho e do usuário ativo (compradores na simulação). A classe *Message* encapsula as informações transmitidas entre compradores e vendedores, podendo ser utilizada em comunicações nos dois sentidos. Implementa acesso aos valores dos itens de contrato propostos e acordados entre as partes, e podem ser de dois tipos: *Propose* e *Counter_Propose*.

Classes do Sistema

O diagrama abaixo apresenta de forma geral as dependências entre as classes descritas nos pacotes do sistema.

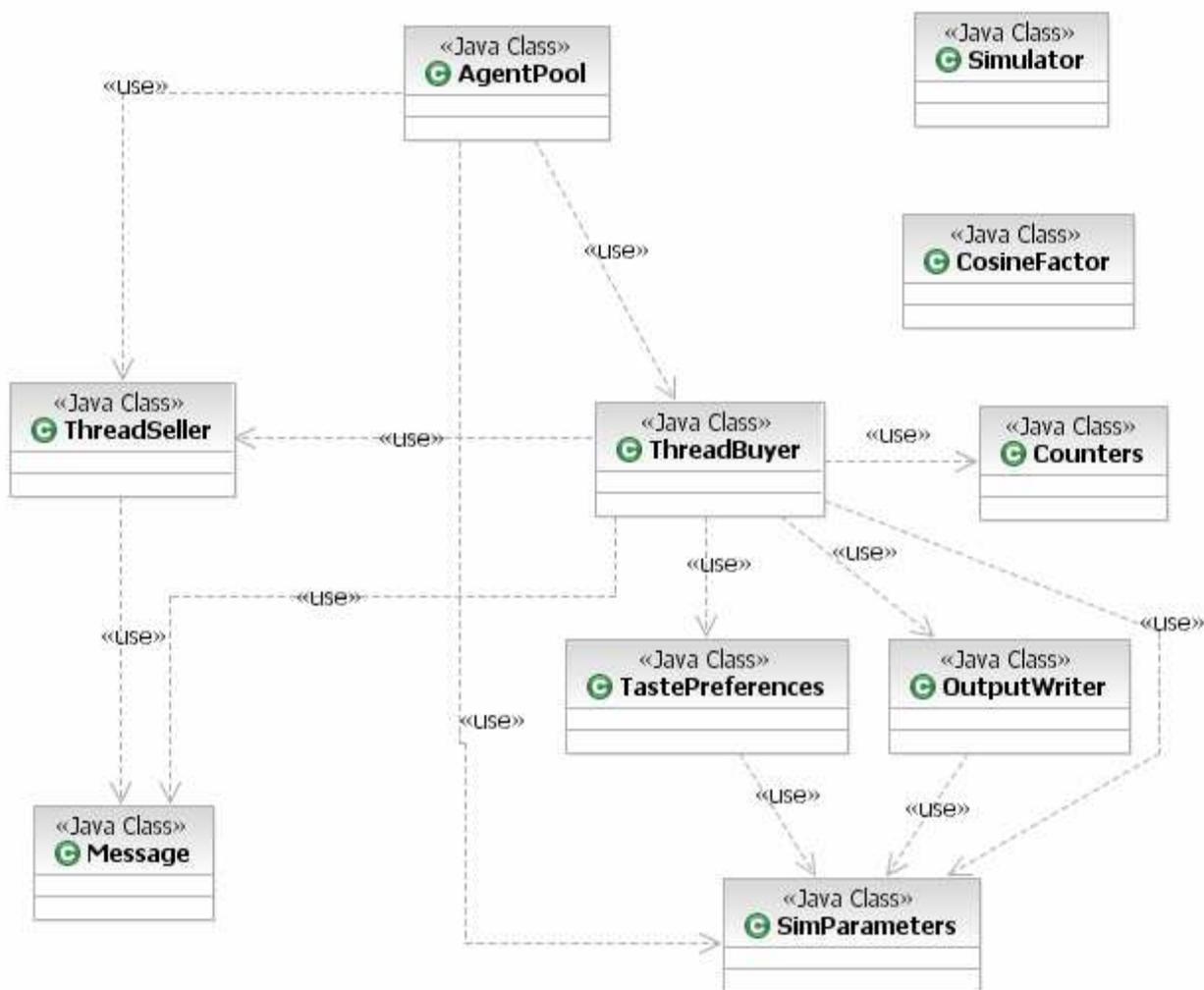


Figura B.8 – Classes da simulação (notação UML 2.0)

Diagramas de Seqüência (Modelo de Design)

Simulação do sistema

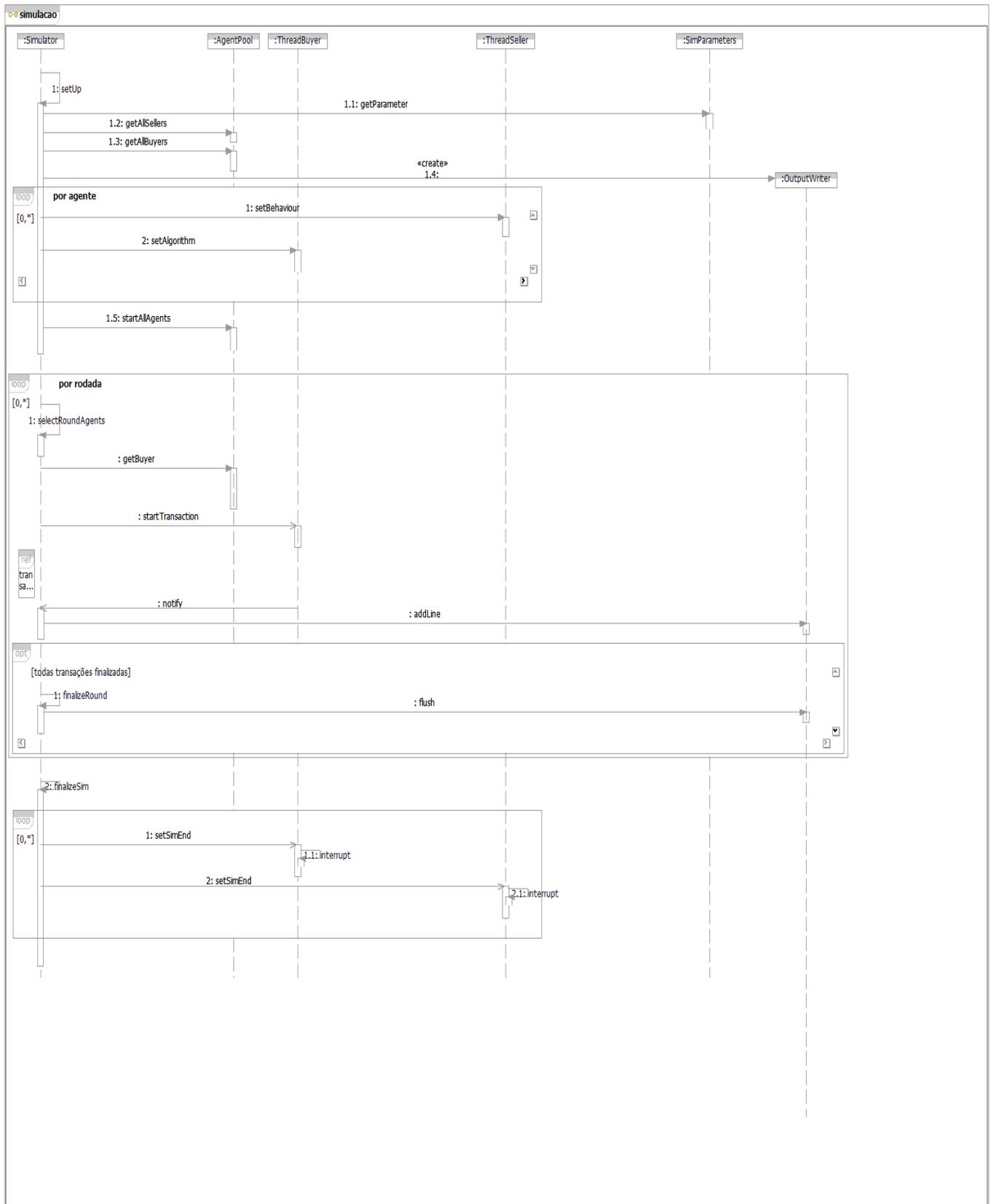


Figura B.9 – Diagrama de seqüência da simulação (notação UML 2.0)

Transações entre Compradores e Vendedores

Interação iniciada a partir da mensagem *startTransaction* do diagrama de seqüência *Simulações do Sistema*.

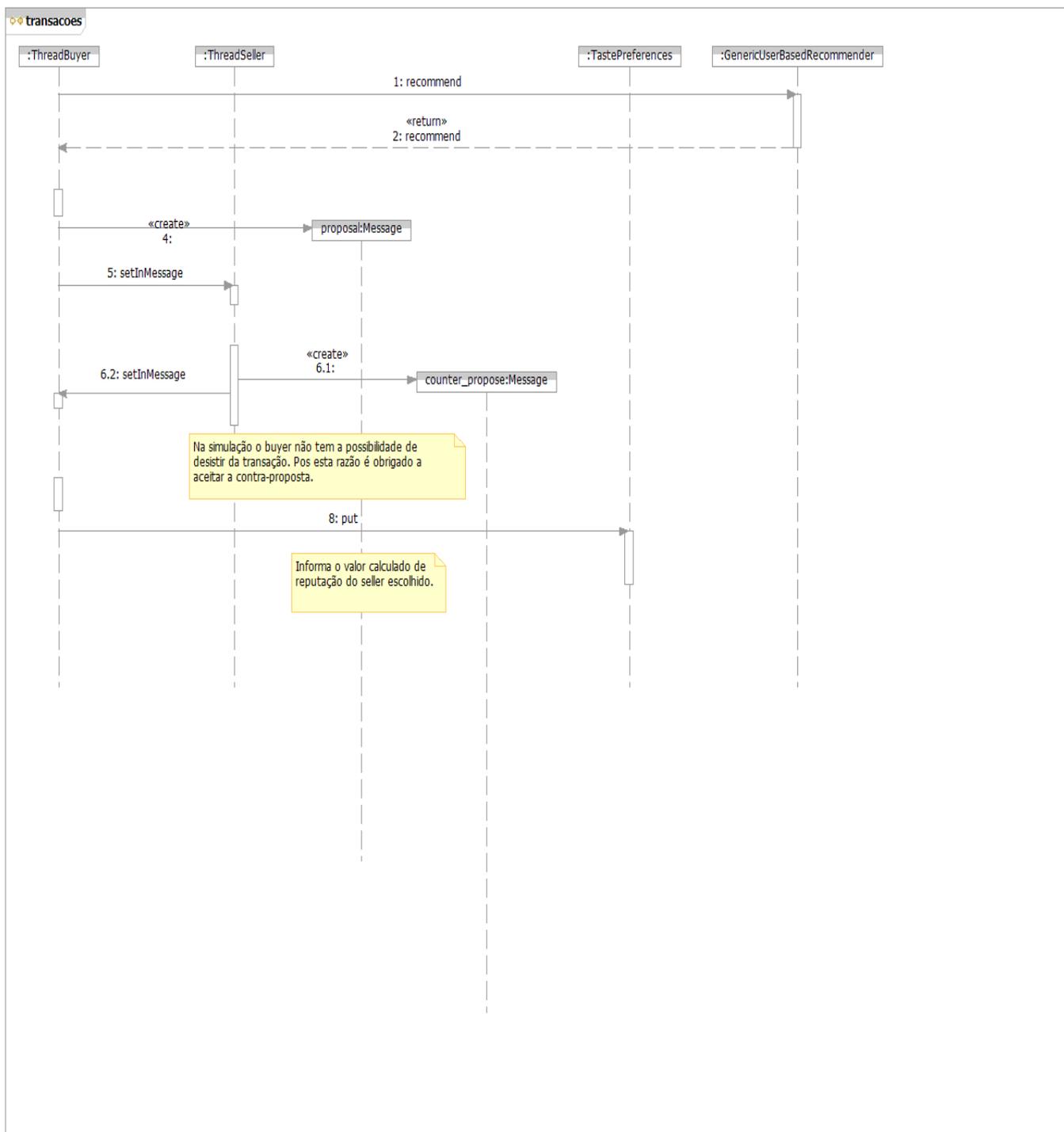


Figura B.10 – Diagrama de seqüência da transação entre comprador e vendedor (notação UML 2.0)

Interface de usuário do sistema

Em razão da própria simulação simular os usuários do sistema, não foram projetadas interfaces de usuário.

Por esta razão não apresentamos *design* de telas nem *Story Boards* da IU.

Arquivo de parâmetros da simulação

São apresentados abaixo dois exemplos de arquivos de parâmetros de simulação, cada qual referente a um tipo de teste.

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE properties SYSTEM "http://java.sun.com/dtd/properties.dtd">
<!--
  Document      : simulation_parameters.xml
  Created on    : 4 de Setembro de 2005, 20:08
  Author       : Alexandre Lopes
  Description:
    Cenário I - teste comparativo 1 - método de recomendação FCA1
-->
<properties>
<entry key="name">Minha simulação</entry>
<entry key="qtd_rounds">45</entry>
<entry key="qtd_train_rounds">25</entry>
<entry key="qtd_transacoes_round">8</entry>
<entry key="qtd_buyers">8</entry>
<entry key="qtd_sellers">32</entry>
<entry key="qtd_neutral_behaviour">3</entry>
<entry key="qtd_bad_behaviour">13</entry>
<entry key="qtd_good_behaviour">16</entry>
<entry key="qtd_evaluation_items">2</entry>
<entry key="price_min_value">10.0</entry>
<entry key="price_max_value">50.0</entry>
<entry key="scale_max">6</entry>
<entry key="price_agreed_value">30.0</entry>
<entry key="quality_min_value">1</entry>
<entry key="quality_max_value">5</entry>
<entry key="quality_agreed_value">3</entry>
<entry key="qtd_neighbours">30</entry>
<entry key="qtd_recommendations">5</entry>
<entry key="perc_time_defect">0.60</entry>
<entry key="price_increase_factor">0.6666</entry>
<entry key="price_decrease_factor">0.6666</entry>
<entry key="quality_increase_factor">0.6666</entry>
<entry key="quality_decrease_factor">0.6666</entry>
<entry key="initial_cash">5000.00</entry>
<entry key="price_factor">10</entry>
<entry key="output_file">./simulator/simulator/output/tst</entry>
<entry key="scenario">ScenarioI</entry>
<entry key="test_option">4</entry> <!-- Método de recomendação FCA1 -->
</properties>
```

```

<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE properties SYSTEM "http://java.sun.com/dtd/properties.dtd">
<!--
  Document      : simulation_parameters.xml
  Created on    : 4 de Setembro de 2005, 20:10
  Author       : Alexandre Lopes
  Description:
      Cenário I - teste comparativo 2 - método de recomendação FCA1
-->
<properties>
<entry key="name">Minha simulação</entry>
<entry key="qtd_rounds">114</entry>
<entry key="qtd_train_rounds">81</entry>
<entry key="qtd_transacoes_round">13</entry>
<entry key="qtd_buyers">13</entry>
<entry key="qtd_sellers">52</entry>
<entry key="qtd_neutral_behaviour">5</entry>
<entry key="qtd_bad_behaviour">21</entry>
<entry key="qtd_good_behaviour">26</entry>
<entry key="qtd_evaluation_items">2</entry>
<entry key="price_min_value">10.0</entry>
<entry key="price_max_value">50.0</entry>
<entry key="scale_max">6</entry>
<entry key="price_agreed_value">30.0</entry>
<entry key="quality_min_value">1</entry>
<entry key="quality_max_value">5</entry>
<entry key="quality_agreed_value">3</entry>
<entry key="qtd_neighbours">30</entry>
<entry key="qtd_recommendations">5</entry>
<entry key="perc_time_defect">0.60</entry>
<entry key="price_increase_factor">0.6666</entry>
<entry key="price_decrease_factor">0.6666</entry>
<entry key="quality_increase_factor">0.6666</entry>
<entry key="quality_decrease_factor">0.6666</entry>
<entry key="initial_cash">5000.00</entry>
<entry key="price_factor">10</entry>
<entry key="output_file">./simulator/simulator/output/tst</entry>
<entry key="scenario">ScenarioI</entry>
<entry key="test_option">4</entry> <!-- Método de recomendação FCA1 -->
</properties>

```