

1

INTRODUÇÃO

1.1

Motivação

Em um mercado competitivo como o atual, existem alguns fatores que fazem a diferença e que tornam uma empresa líder de mercado em detrimento de outras. Se, no início da internet, as primeiras lojas virtuais puderam prosperar praticamente sem concorrentes, hoje, elas devem competir com gigantes virtuais por fatias de mercado. É bem conhecido o fato de que consumidores online são impacientes e a diferença entre fazer o cliente comprar em sua loja ou visitar a próxima está muitas vezes em fazer que a melhor oferta esteja na frente do cliente antes que ele abandone a página de internet desta loja.

Se, por um lado, ofertar o produto certo ao consumidor é a chave para o sucesso na venda, muitos dos sistemas existentes para recomendação aplicados ao varejo virtual são pouco eficazes no que diz respeito a apresentar ao usuário os itens que lhe interessam dentro da vasta quantidade de itens disponíveis. Sistemas recomendadores baseados em filtros colaborativos e de informação [1,2], que são agentes que usam informação de comportamento ou preferência para filtrar alternativas e propor recomendações ao usuário [3], poderiam trazer a solução. Porém, tais sistemas requerem informação sobre as preferências dos visitantes, como avaliações de produtos, preferências implícitas inferidas de comportamentos do consumidor frente ao navegador ou histórico de compras.

Os Filtros colaborativos apresentam problemas consideráveis com relação ao aumento da escala e da esparsidade dos bancos de dados sobre os quais eles atuam. Possuem também problemas no que se relaciona a existência de itens e usuários novos, para os quais não existem avaliações sobre as quais o filtro possa atuar.

Diversos tipos diferentes de filtros colaborativos foram criados para tentar contornar tais problemas, incluindo filtros baseados em memória e baseados em modelo e híbridos. Atualmente, os filtros baseados em memória têm sido

aplicados comercialmente para recomendações em lojas virtuais como a Amazon. Existe uma ampla área de estudos para aperfeiçoar filtros colaborativos na qual a presente tese se insere.

Anteriores aos filtros colaborativos, os sistemas de recomendação baseados em regras de associação, que utilizavam o algoritmo Apriori [4], já possuíam os mesmos problemas de escala. Tais algoritmos geram recomendações de baixa acurácia a elevados custos computacionais, ainda que continuem em uso em diversas lojas varejistas.

Recentemente, sistemas baseados em conteúdo buscam padrões em textos e dados dos usuários e de itens que permitam a recomendação [5]. Tais algoritmos têm trazido bons resultados porém necessitam da existência de dados prévios dos usuários e itens para funcionarem e podem apresentar problemas de super-especialização. Tais sistemas baseados em conteúdo têm sido utilizado hibridizados com filtros colaborativos para melhorar sua performance.

Por outro lado, desde a década de 50, as empresas empregam as práticas de Peter Drucker [6] de marketing para definir as características de produtos com objetivo de aumentar seu potencial de venda a segmentos específicos de usuários. As técnicas empregadas em marketing são basicamente ignoradas pelos filtros colaborativos e outros sistemas de recomendação. Por outro lado, Hosseinpour [7] criou em 2008 um método de recomendação baseado em conteúdo (que utiliza números fuzzy sobre especificações de itens) que possui paralelos com conceitos de posicionamento de marketing, apesar de este paralelo não ter sido explicitado pelo próprio Hosseinpour em seu artigo. Tais paralelos surgem da definição de “componentes” a partir de especificações de produtos. Tais “componentes”, assim como os posicionamentos, podem ser interpretados como desejos ou interesses de um usuário sobre determinados tipos de itens.

Tendo em vista a elevada relevância que o mercado varejista tem para qualquer país e o fato de ainda existirem problemas a serem resolvidos nos sistemas de recomendação, tanto baseados em filtragem colaborativa quanto os baseados em conteúdo, existe uma grande motivação para estudar esse assunto. A motivação deste estudo é, portanto, desenvolver um algoritmo capaz de gerar recomendações de elevada acurácia para empresas varejistas online, aproveitando conceitos de marketing com uso de algoritmos de filtragem colaborativa e

algoritmos baseados em conteúdo de forma híbrida, buscando obter sinergia entre os mesmos na resolução de suas fraquezas individuais acima relacionadas.

1.2

Objetivos do Trabalho

A proposta deste trabalho é desenvolver um novo algoritmo de recomendação que explore ao máximo a união entre as tecnologias presentes de recomendação (sistemas baseados em conteúdo e filtros colaborativos) e de conceitos oriundos da teoria de marketing, para que cada algoritmo compense as fraquezas dos outros.

As fraquezas que devem ser solucionadas com este novo algoritmo, para o caso dos filtros colaborativos, se traduzem na dificuldade do algoritmo de gerar recomendações de elevada acurácia em bases de dados grandes e esparsas, bem como o problema de recomendação no caso da existência do banco de dados de itens ou usuários novos. Quanto aos algoritmos baseados em conteúdo, o novo algoritmo deve ser capaz de utilizar apenas uma quantidade mínima de dados do usuário, uma vez que estes são difíceis de serem obtidos em lojas virtuais por motivos de privacidade.

O modelo do novo algoritmo inclui o uso de um filtro colaborativo para definição da categoria de item a ser recomendada, seguido por um algoritmo baseado em conteúdo para definição do item dentro da categoria escolhida que melhor satisfaz as necessidades de compra do usuário. O algoritmo baseado em conteúdo utiliza números fuzzy para gerar posicionamentos de marketing baseados nas especificações técnicas dos itens. O algoritmo gerado, portanto, é um sistema híbrido.

O software desenvolvido com os algoritmos pesquisados permitirá o teste do sistema final em lojas virtuais, de modo a comprovar a sua eficiência quando comparado a um algoritmo de filtragem colaborativo comumente usado no mercado para recomendação.

O modelo proposto nesse trabalho combinam os paradigmas de modelagem por números fuzzy, conceitos de posicionamento de produtos oriundos da teoria de marketing e filtros colaborativos de produtos. O sistema resultante é capaz de

oferecer resultados iniciais para produtos e usuários novos e se adaptar ao longo do uso aos padrões de consumo existentes.

1.3

Contribuições da Tese

A principal contribuição desse trabalho é:

- *Desenvolvimento de um filtro colaborativo híbrido* → Modelo composto por um filtro colaborativo categórico seguido por um filtro baseado em números fuzzy. Tal filtro colaborativo híbrido foi capaz de inserir o conceito de posicionamento de marketing nas decisões de recomendação de itens para usuários, tendo invariância à esparsidade da base de dados.

Outras contribuições desse trabalho são:

- *Interpretação do método de recomendação Hosseinpour para marketing* → Tal interpretação permitiu que as especificações de produtos fossem interpretadas em posicionamentos de produtos utilizados em teoria de marketing, o que permitiu não só a criação do algoritmo híbrido proposto nesta tese, como a criação da base de dados sintética com o qual o algoritmo foi testado.
- *Base de Dados Sintética* → Existe uma grande dificuldade no estudo de sistemas de recomendação em obter bases de dados para avaliar o desempenho de algoritmos. A interpretação do método de recomendação Hosseinpour para marketing permitiu a criação de uma base sintética cujo comportamento artificial de compra dos consumidores segue regras realísticas e replicáveis, sendo o algoritmo para criação desta base de dados uma contribuição importante para esta área do conhecimento.

1.4

Estrutura da Tese

O restante desta tese está dividido conforme descrito a seguir. No capítulo 2, será apresentada a teoria de marketing utilizada para empresas ao criar novos

produtos e serviços com base em nichos de mercado. Os capítulos 3 e 4 apresentam o estado da arte. O capítulo 3 oferece uma visão geral de algoritmos de filtragem colaborativa utilizados para recomendação de itens. O capítulo 4 apresenta outros algoritmos de recomendação, incluindo um sistema de recomendação baseado em números fuzzy que será utilizado posteriormente para mesclar as técnicas de marketing com sistemas de recomendação baseados em conteúdo.

O capítulo 5 apresenta o modelo proposto, que traça um paralelo com o posicionamento de marketing e um algoritmo de recomendação baseado em números fuzzy apresentado no capítulo anterior, e define as escolhas feitas em cada algoritmo específico para a composição do algoritmo final de recomendação. Serão explicitadas as razões para tais escolhas e a forma como esses algoritmos são conectados.

No capítulo 6 é descrito um sistema de testes para analisar a capacidade do algoritmo em se adaptar a diversos padrões de compra e interesses do consumidor. Este sistema de testes é composto por uma base de dados artificial, cujo objetivo é testar o algoritmo criado em comparação com um filtro colaborativo amplamente utilizado em lojas virtuais sob o ponto de vista de diversos valores de esparsidade. Também é utilizada uma base de dados real amplamente utilizada para comparação de algoritmos de recomendação para analisar sua eficácia em recomendar itens.

O capítulo 7 apresenta as conclusões e sugestões para trabalhos futuros.