Pontifícia Universidade Católica
do Rio de Janeiro

Miguel Angelo Gaspar Pinto

Sistema Híbrido de Recomendação de Produtos com Uso de Filtros Colaborativos e Números Fuzzy

Tese de Doutorado

Tese apresentada ao programa de Pós-Graduação em Engenharia Elétrica do Departamento de Engenharia Elétrica da PUC-Rio como parte dos requisitos parciais para obtenção do título de Doutor em Engenharia Elétrica.

Orientador: Prof. Ricardo Tanscheit
Co-orientadora: Profa. Marley Maria Bernardes Rebuzzi Vellasco

Rio de Janeiro

Setembro de 2011

PONTIFÍCIA UNIVERSIDADE CATÓLICA
DO RIO DE JANEIRO

Miguel Angelo Gaspar Pinto

Sistema Híbrido de Recomendação de Produtos com uso de Filtros Colaborativos e Números Fuzzy

Tese apresentada como requisito parcial para obtenção do grau de Doutor pelo Programa de Pós-Graduação em Engenharia Elétrica do Departamento de Engenharia Elétrica do Centro Técnico Científico da PUC-Rio. Aprovada pela Comissão Examinadora abaixo assinada.

Prof. Ricardo Tanscheit
Orientador
Departamento de Engenharia Elétrica – PUC-Rio

Profa. Marley Maria Bernardes Rebuzzi Vellasco
Co-Orientadora
Departamento de Engenharia Elétrica – PUC-Rio

Profa. Karla Tereza Figueiredo Leite
UEZO

Prof. Jorge Luís Machado do Amaral
UERJ

Prof. Christian Nunes Aranha
Cortex

Prof. Juan Guillermo Lazo Lazo
ICA/DEE/PUC-Rio

Prof. Roxana Jimenez Contreras
Departamento de Engenharia Elétrica – PUC-Rio

Prof. Emmanuel Piseces Lopes Passos
Graall Consultoria Empresarial Ltda

Prof. José Eugenio Leal
Coordenador Setorial do Centro
Técnico Científico

Rio de Janeiro, 27 de setembro de 2011

Miguel Angelo Gaspar Pinto

Graduou-se em Engenharia de Controle e Automação (Pontifícia Universidade Católica) em 2004. Trabalhou por meio ano na área de controle antes de começar mestrado na área de Processamento de Sinais e Controle na Pós-Graduação da PUC-Rio. Participou de congressos na área de controle e inteligência artificial. Atualmente trabalha em projetos de inovação em empresa própria. Suas áreas de interesse abrangem robótica, controle de sistemas, visão computacional e inteligência artificial.

Ficha Catalográfica

Pinto, Miguel Angelo Gaspar

Sistema híbrido de recomendação de produtos com uso de filtros colaborativos e números fuzzy / Miguel Angelo Gaspar Pinto; orientador: Ricardo Tanscheit; co-orientadora: Marley Maria Bernardes Rebuzzi Vellasco 2011.

129 f. ; 30 cm

Tese (doutorado) – Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Engenharia Elétrica, 2011.

Inclui bibliografia

1. Engenharia elétrica – Teses. 2. Marketing. 3. Números fuzzy. 4. Filtros colaborativos. 5. Recomendação. 6. Algoritmos baseados em conteúdo. I. Tanscheit, Ricardo. II. Vellasco, Marley Maria Bernardes Rebuzzi. III. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Engenharia Elétrica. IV. Título.

CDD: 621.3

Dedico esta tese ao meu pai, que não me deixou desistir de terminá-la apesar das pressões para o contrário. Dedico também a minha mãe que faleceu e que espero esteja orgulhosa por mais este feito.

Agradecimentos

Agradeço aos meus orientadores pela enorme paciência que tiveram. Agradeço também a minha esposa pela compreensão e por aceitar os vários fins de semana que deixei de aproveitar com ela para completar este feito.

Resumo

Pinto, Miguel Ângelo Gaspar; Tanscheit, Ricardo (Orientador); Vellasco, Marley Maria Bernardes Rebuzzi (Co-orientadora). **Sistema Híbrido de Recomendação de Produtos com uso de Filtros Colaborativos e Números Fuzzy**. Rio de Janeiro, 2011. 129p. Tese de Doutorado. Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro.

O varejo virtual tem sido um importante setor para dinamização da economia, cujo valor das transações em 2010 ficou em torno de R$10,6 bilhões. As lojas nesse segmento não possuem restrição de clientes ou de estoque, porém possuem consumidores pouco pacientes com várias outras lojas a sua disposição, sendo necessário que o item de seu interesse seja encontrado visível rapidamente. Buscando resolver este problema, foram desenvolvidos algoritmos de recomendação capazes de gerar listagens de produtos que fossem direcionados ao usuário. Os algoritmos de filtragem colaborativa são amplamente usados no varejo virtual, porém eles apresentam problemas devido a escala e esparsidade do banco de dados. Algoritmos baseados em conteúdo podem apresentar menor sensibilidade ao tamanho da base de dados, porém sua efetividade depende da existência de dados de usuários que comumente não estão presentes. Nesta tese, propõe-se um algoritmo híbrido que utiliza tanto a filtragem colaborativa quanto um algoritmo baseado em conteúdo para permitir boas recomendações em bases de dados esparsas e de grande porte. O algoritmo baseado em conteúdo faz uso de números fuzzy e técnicas de marketing para guiar sua recomendação apenas com base nos itens comprados pelo usuário, sem necessidade de quaisquer outros dados pessoais do usuário. O algoritmo proposto foi testado em bases de dados sintética e real, sendo comparado com um filtro colaborativo padrão para avaliar seu desempenho.Os resultados obtidos demonstram que o algoritmo híbrido proposto apresentou um desempenho superior ao do filtro colaborativo padrão em ambas as base de dados, apresentando invariância à esparsidade da base de dados.

Palavras-chave

Marketing; Números Fuzzy; Filtros Colaborativos; Recomendação, Algoritmos baseados em conteúdo.

Abstract

Pinto, Miguel Ângelo Gaspar; Tanscheit, Ricardo (Advisor); Vellasco, Marley Maria Bernardes Rebuzzi (Co-advisor). **Hybrid Recommendation System Based on Collaborative Filtering and Fuzzy Numbers**. Rio de Janeiro, 2011. 129p. PhD Thesis – Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro.

The virtual retail has been an important sector at Brazilian economy, being a $6.23 billion market in 2010, having 30% expansion on that period. The companies in such segment don't have client or product restrictions due to physical limitations. On the other hand, the consumers of this kind of retail have several options to buy and little patience to keep searching on the same website. The companies need to define which item will be shown to the consumer before he leaves for the next competitor. Several recommendation algorithms were developed to generate products list directed to the consumer. Nowadays the algorithms for collaborative filtering are well spread in virtual retail, but they have problems caused exactly by the huge quantity of data that exist on virtual retail. Content based algorithms are less sensitive to the size of the database, but their effectiveness depends on the existence of user data, which usually are not available. This thesis proposes a hybrid algorithm which uses both collaborative filtering and a content based algorithm to allow recommendations in huge sparse databases. The content base algorithm uses fuzzy numbers and marketing techniques to guide the recommendation using only the items brought by the user, without the need for further personal data from the consumer. The proposed algorithm was tested in both artificial and real databases, compared with a benchmark collaborative filter. The collected results show that the proposed hybrid algorithm provides superior performance than the benchmark collaborative filter in both databases, generating good results and presenting sparsity invariance. The proposed algorithm also solves problems of initialization, neighborhood transitivity and in cases when new users or items are inserted on database.

Keywords

Marketing; Fuzzy Numbers; Collaborative Filters; Recommendation, Content Based Algorithms

Sumário

Lista de figuras

Lista de tabelas

Lista de quadros