Universidade Federal de Pernambuco

Centro de Informática

Pós-Graduação e Ciência da Computação

Clayton Rodrigues da Silva

Investigação da Combinação de Filtragem Colaborativa e Recomendação Baseada em Confiança Através de Medidas de Esparsidade

Recife, PE

2017

Universidade Federal de Pernambuco

Centro de Informática

Pós-Graduação em Ciência da Computação

Investigação da Combinação de Filtragem Colaborativa e Recomendação Baseada em Confiança Através de Medidas de Esparsidade

**Dissertação de Mestrado** apresentada ao Centro de Informática da Universidade Federal de Pernambuco por **Clayton Rodrigues da Silva**, sob a orientação do **Prof. Dr. Robson do Nascimento Fidalgo**, como parte dos requisitos para obtenção do título de Mestre em Ciência da Computação.

Recife, PE

2017

Dissertação de Mestrado Profissional apresentada por **Clayton Rodrigues da Silva** à Pós-Graduação em Ciência da Computação do Centro de Informática da Universidade Federal de Pernambuco, sob o título, “**Investigação da Combinação de Filtragem Colaborativa e Recomendação Baseada em Confiança Através de Medidas de Esparsidade**”, orientada pelo Professor Ricardo Bastos Cavalcante Prudêncio e aprovada pela Banca Examinadora formada pelos professores:

Prof. Ricardo Bastos Cavalcante Prudêncio

CIn/UFPE

Prof. Renato Fernandes Corrêa

CAC/Departamento de Ciência da Informação/UFPE

Prof. Cleber Zanchettin

CIn/UFPE

Visto e permitida a impressão.

Recife, 06 de Agosto de 2015.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Profª Edna Natividade da Silva Barros**

Coordenadora de Pós-Graduação em Ciência da Computação

da Universidade Federal de Pernambuco

Dedico este trabalho a minha família em especial a minha avó Eunice que já faleceu, minha mãe Martha que sempre me incentivou nos estudos, a minha irmã Iris e minhas sobrinhas Iasmin e Iara.

Agradecimentos

Gostaria de agradecer primeiramente a Deus, por ter me guiado até este momento com saúde e por ter colocado pessoas maravilhosas na minha vida.

.

Resumo

Sistemas de recomendação têm desempenhado um papel importante em diferentes contextos de aplicação (e.g recomendação de produtos, filmes, músicas, livros, dentre outros). Eles automaticamente sugerem a cada usuário itens que podem ser relevantes, evitando que o usuário tenha que analisar uma quantidade gigantesca de itens para realizar sua escolha.

Palavras-Chave: Sistemas de Recomendação, Filtragem Colaborativa, Fatoração de Matriz, Sistemas de Recomendação Baseados em Confiança, Medidas de Esparsidade

Abstract

Recommender systems have played an important role in different application contexts (e.g recommendation of products, movies, music, books, among others).

Keywords: Recommender Systems, Collaborative Filtering, Matrix Factorization, Trust-Based Recommendation Systems, Sparsity Measures

Lista de Figuras

[Figura 2‑1 - FC baseada em memória 13](#_Toc468925691)

[**Figura 3‑1**: Proposta Arquitetural 16](#_Toc468925692)

[**Figura 3‑2**: Criação de uma hierarquia de conceitos na base multidimensional 16](#_Toc468925693)

Lista de Tabelas

[Tabela 2.1 - Exemplo de matriz de usuários x itens 13](#_Toc468925719)

Lista das Abreviaturas/Acrônimos

MDX Multi-Dimensional Expression

OLAP Online Analytical Processing

DW Data Warehouse

SQL Structured Query Language

DL Description Logic

Sumário

[Investigação da Combinação de Filtragem Colaborativa e Recomendação Baseada em Confiança Através de Medidas de Esparsidade - 1 -](#_Toc468925738)

[Agradecimentos - 5 -](#_Toc468925739)

[Resumo - 5 -](#_Toc468925740)

[Abstract - 7 -](#_Toc468925741)

[Lista de Figuras - 7 -](#_Toc468925742)

[Lista de Tabelas - 8 -](#_Toc468925743)

[Lista das Abreviaturas/Acrônimos - 8 -](#_Toc468925744)

[Sumário - 9 -](#_Toc468925745)

[1 Introdução 12](#_Toc468925746)

[2 Conceitos Básicos 13](#_Toc468925747)

[2.1. Introdução aos Sistemas de Recomendação 13](#_Toc468925748)

[3 Trabalhos Relacionados 14](#_Toc468925749)

[3.1. Características para avaliação 14](#_Toc468925750)

[3.2. Trabalhos Acadêmicos 15](#_Toc468925751)

[3.2.1. Anderlik et. al. 15](#_Toc468925752)

[3.2.2. Neumayr et. al. 17](#_Toc468925753)

[4 OLAP e Ontologia 18](#_Toc468925754)

[5 Avaliação dos trabalhos 19](#_Toc468925755)

[6 Conclusões 20](#_Toc468925756)

[Referências Bibliográfica 21](#_Toc468925757)

# Introdução

Atualmente, com a grande quantidade de produtos, serviços, livros, revistas, trabalhos científicos, músicas, amigos em redes sociais, e outras infinidades de itens que os usuários se deparam no uso da internet, sistemas de recomendação têm se tornado uma ferramenta importante para fornecer recomendações mais personalizadas a esses usuários.

# Conceitos Básicos

Este capítulo está organizado do seguinte modo. A seção 2.1 fará uma breve introdução sobre sistemas de recomendação e as principais abordagens utilizadas.

## Introdução aos Sistemas de Recomendação

Em um cenário típico, os dados deste sistema são representados através de uma matriz, onde usuários estão dispostos em linhas e itens dispostos em colunas. No caso de um sistema de recomendação de filmes, uma matriz de usuários x itens utilizando a escala mencionada seria semelhante a da tabela 3.1.

Tabela 2.1 - Exemplo de matriz de usuários x itens

|  |  |  |  |
| --- | --- | --- | --- |
|  | Os Vingadores | Shrek | O Poderoso Chefão |
| Marcos | 4 | 2 |  |
| Silvia |  | 3 | 1 |
| Pedro | ? | 4 | 2 |

Na tabela 2.1, Pedro seria o usuário ativo para o qual nós gostaríamos de predizer a nota para o filme “Os Vingadores”. Possivelmente, podem existir alguns ou vários valores faltantes (matriz muito esparsa) em uma matriz deste tipo. eles avaliaram.

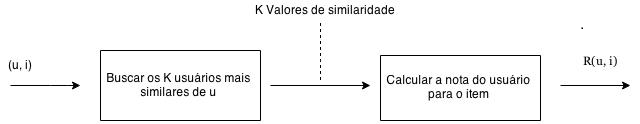


Figura 2‑1 - FC baseada em memória

# Trabalhos Relacionados

Neste capítulo será feita uma análise dos principais trabalhos relacionados ao objeto de pesquisa desta dissertação e, quando for o caso, dos respectivos protótipos utilizados como prova de conceito. Para realizar esta análise, foram selecionados trabalhos que façam uso de ontologias para enriquecimento semântico de consultas OLAP.

## Características para avaliação

Procurando efetuar uma análise equânime dos trabalhos relacionados, apresentaremos, a seguir, os critérios que nortearam essa avaliação. Para cada trabalho é avaliado se ele cobre (Sim), não cobre (Não) ou cobre parcialmente (Parcial) cada um dos critérios de avaliação.

1. **Uso da linguagem MDX**: analisa se a interação com o Data Warehouse é feita por intermédio da linguagem MDX;
2. **Suporte a múltiplas ontologias**: verifica se o trabalho analisado dá suporte ao uso de múltiplas ontologias nas atividades relacionadas ao enriquecimento semântico de um DW;
3. **Uso de raciocinador**: investiga se há uso de um raciocinador para deduzir informações implícitas na ontologia e no DW;
4. **Slice and Dice**: Examina se o trabalho científico dá suporte às operações Slice and Dice;
5. **Roll Up and Drill Down**: examina se o trabalho científico dá suporte às operações Roll Up and Drill Down;
6. **Uso de base multidimensional**: averigua se a base de dados utilizada é multidimensional;
7. **Integração com outras aplicações**: analisa se o trabalho relacionado pode ser utilizado em conjunto com outras aplicações;
8. **Uso de dimensões semânticas**: verifica a existência, ou não, de uma ontologia sendo utilizada como dimensão semântica;
9. **Mapeamento**:investiga se há algum tipo de mapeamento entre a ontologia e o DW.
10. **Lógica Descritiva**: examina se o trabalho analisado fez uso de consultas usando Lógica Descritiva.

## Trabalhos Acadêmicos

Nesta seção, faz-se uma análise dos trabalhos acadêmicos selecionados. Apesar de não ter sido realizada uma revisão sistemática, a triagem dos trabalhos aqui apresentados foi realizada a partir de pesquisas nas bases de dados IEEE, ACM, Springer, Science Direct e Google Scholar. Esta pesquisa foi feita por meio da seguinte expressão: “Data Warehouse AND Ontology AND OLAP”. De posse dos resultados, uma análise foi realizada a partir do título e do abstract com o intuito de descartar os trabalhos que não fizessem uso de ontologias para o enriquecimento semântico de bases multidimensionais.

### Anderlik et. al.

O trabalho de (Anderlik et. al., 2012) propõe uma abordagem que tem como objetivo principal ilustrar como ontologias de domínio podem ser utilizadas para melhorar a análise de um Data Warehouse. Para isso, o conhecimento do negócio foi modelado em uma ontologia de domínio que, por sua vez, é utilizada para agregação de dados através de uma hierarquia de conceitos nela modelados. Um ponto forte deste trabalho é a integração entre uma ontologia de domínio e a base multidimensional ou mapeamentos entre elas. A justifica para a ausência de mapeamentos reside, segundo os autores, no seguinte argumento: o mapeamento de atributos semânticos para valores não ontológicos ocasiona perda de conhecimento. Outro ponto que merece destaque é a capacidade de executar as operações Drill Down and Roll Up, Slice and Dice em uma hierarquia de conceitos presentes em ontologia de domínio. Convém mencionar que o presente trabalho permite o uso de consultas em DL para a recuperação de informações no DW.

Ao analisar as Figuras 3.1 e 3.2, que serão apresentadas a seguir, e tendo em vista os critérios avaliativos apresentados na seção 3.1, podemos fazer algumas constatações acerca do trabalho hora analisado. A primeira delas é que o trabalho não faz uso da linguagem MDX para interagir com a base de dados multidimensional, e sim a linguagem SQL. Além disso, as informações são persistidas em uma base de dados relacional, conforme apresentado na Figura 3.1. Adicionalmente, o protótipo hora avaliado faz uso de um raciocinador para inferir hierarquias de conceitos modelados na ontologia de domínio, armazenando-as como views materializadas em um Data Warehouse Híbrido (DWH).

|  |  |
| --- | --- |
| **Figura 3‑1**: Proposta Arquitetural  para integração entre OLAP e ontologia | **Figura 3‑2**: Criação de uma hierarquia de conceitos na base multidimensional |

**Fonte:** (Anderlik et. al., 2012)

Por fim, é importante mencionar que o trabalho analisado não permite a integração com outras aplicações bem como não dá suporte ao uso de mais de uma ontologia em sua proposta arquitetural. Com o intuito de sintetizar o atendimento, ou não, dos critérios avaliativos apresentados na seção 3.1 por parte do trabalho hora analisado, apresentamos, a seguir, o Quadro 3.1.

|  |  |
| --- | --- |
| **Característica da Avaliação** | **Sim / Não** |
| Uso da linguagem MDX | Não |
| Suporte a múltiplas ontologias | Não |
| Uso de raciocinador | Sim |
| Slice and Dice | Sim |
| Roll Up and Drill Down | Sim |
| Uso de base multidimensional | Sim |
| Integração com outras aplicações | Não |
| Uso de dimensões semânticas | Sim |
| Mapeamento | Não |
| Lógica Descritiva | Sim |

**Quadro 3‑1**: Avaliação do trabalho proposto por (Anderlik et. al., 2012)

**Fonte**: Autor

### Neumayr et. al.

# OLAP e Ontologia

Neste capítulo, iremos responder as seguintes perguntas: O que foi avaliado? Que conjunto de dados foi utilizado? Como foram comparados os resultados? A primeira pergunta será respondida na seção 5.3, onde serão descritos os algoritmos comparados. O conjunto de dados será apresentado na seção 5.1 e as métricas utilizadas para comparação dos algoritmos serão apresentadas na seção 5.2.

# Avaliação dos trabalhos

Neste capítulo serão apresentados os resultados obtidos a partir da proposta de algoritmo Híbrido com as medidas de esparsidade propostas (seção 4.3) e os métodos base que compõem a arquitetura deste trabalho (seção 5.3).

# Conclusões

Os sistemas de recomendação têm sido uma ferramenta importante de sugestão informações para usuários. Como descrito no capítulo 2, FC tem se tornado a abordagem mais popular para sistemas de recomendação, embora sofra de algumas fraquezas intrínsecas, como por exemplo, os usuários com poucas avaliações. Para amenizar esse e outros problemas em FC tradicional, surgiram estratégias que utilizam as informações obtidas das redes sociais sobre confiança entre os usuários.

# Referências Bibliográfica

ANDERLIK, Stefan; NEUMAYR, Bernd; SCHREFL, Michael. Using domain ontologies as semantic dimensions in data warehouses. In: Conceptual Modeling. Springer Berlin Heidelberg, 2012. p. 88-101.

NEUMAYR, Bernd; SCHÜTZ, Christoph; SCHREFL, Michael. Semantic enrichment of OLAP cubes: Multi-dimensional ontologies and their representation in SQL and OWL. In: On the Move to Meaningful Internet Systems: OTM 2013 Conferences. Springer Berlin Heidelberg, 2013. p. 624-641.