Chapter VIII
Evaluating the Construction of Domain Ontologies for Recommender Systems Based on Texts

Stanley Loh
Catholic University of Pelotas and Lutheran University of Brazil, Brazil

Daniel Lichtnow
Catholic University of Pelotas, Brazil

Thyago Borges
Catholic University of Pelotas, Brazil

Gustavo Piltcher
Catholic University of Pelotas, Brazil

ABSTRACT

This chapter investigates different aspects in the construction of a domain ontology to a content-based recommender system. The recommender systems suggests textual electronic documents from a digital library, based on documents read by the users and based on textual messages posted in electronic discussions through a Web chat. The domain ontology is used to represent the user's interest and the content of the documents. In this context, the ontology is composed by a hierarchy of concepts and keywords. Each concept has a vector of keywords with weights associated. Keywords are used to identify the content of the texts (documents and messages), through the application of text mining techniques. The chapter discusses different approaches for constructing the domain ontology, including the use of text mining software tools for supervised learning, the interference of domain experts in the engineering process and the use of a normalization step.

INTRODUCTION

Recommender systems are an emerging technology recently used by marketing departments and vendors, that utilizes data mining and machine learning techniques to analyze behavior and preferences of users or customers and that suggests new offers.

A content-based recommender system utilizes the content of items to generate recommendations to users. Items may be products, services, objects, Websites, people, etc. According to Burke (2002), content-based recommendation is an evolution of information filtering research. In the content-based approach, the recommender system learns a profile of the user's interests based on the features present in objects associated to the user and recommends other items with similar features. When the content of an item matches the interest of a user, this item is recommended to that user.

Some works have reported improvements in retrieval tasks using ontologies. Gauch et al., (2003) argued that "*one increasingly popular way to structure information is through the use of ontologies, or graphs of concepts.*" Labrou and Finin (1999) use categories from *Yahoo!* as an ontology to describe content and features of documents. Middleton et al., (2003) associate concepts from an ontology to users' profiles and to documents.

In a recommender system, an ontology may be used to identify and represent the content of the items and the interest (profile) of the users. For example, supermarkets can use ontologies to classify products in sections and brands and the interest of the user may be composed by sections and brands of the items bought by the user. Then a recommender system may suggest to the user other items with the same brand or within the same section.

In this work, we consider an ontology as a formal and explicit definition of concepts (classes or categories) and their attributes and relations (Noy & McGuinness, 2003). A domain ontology is a description of "things" that exist or can exist in a domain (Sowa, 2002) and contains the vocabulary related to the domain (Guarino, 1998).

When dealing with items that have textual descriptions, the ontology may manage keywords that are used to represent the content of the items or user's profile. Textual descriptions may appear as product information (i.e., marketing slogans, technical specifications) or as content of the item (i.e., textual documents, Websites, e-mail messages).

The process of constructing such a domain ontology is important to the recommender system generate better recommendations. There are many approaches for constructing ontologies. This chapter deals with the special case of creating domain ontologies based on keywords or textual characteristics. Ontological engineering techniques must consider the textual content of the items. For this purpose, text mining techniques may be used in a supervised learning process to automate part of the construction process, minimizing the charge on the engineer.

The chapter presents the approaches based on a real recommender system that suggests electronic documents from a digital library based on documents read by users and based on textual messages sent during electronic discussions in a Web chat.

GENERAL EXPLANATION OF THE RECOMMENDER SYSTEM

This section presents the recommender system used as case study. The goal of the system is to provide people with useful information during a collaboration session. To do that, the system analyzes textual messages sent by users when interacting in a private Web chat, identifies topics (subjects/themes/concepts) inside the messages and recommends items cataloged in a private digital library, previously classified in the same

Related Content

An Efficient Association Rule Mining-Based Spatial Keyword Index
Lianyin Jia, Haotian Tang, Mengjuan Li, Bingxin Zhao, Shoulin Weiand Haihe Zhou (2023). *International Journal of Data Warehousing and Mining (pp. 1-19).*
www.irma-international.org/article/an-efficient-association-rule-mining-based-spatial-keyword-index/316161

Referential Horizontal Partitioning Selection Problem in Data Warehouses: Hardness Study and Selection Algorithms
Ladjel Bellatreche, Kamel Boukhalfa, Pascal Richardand Komla Yamavo Woameno (2009). *International Journal of Data Warehousing and Mining (pp. 1-23).*
www.irma-international.org/article/referential-horizontal-partitioning-selection-problem/37402

A BPMN-Based Design and Maintenance Framework for ETL Processes
Zineb El Akkaoui, Esteban Zimányi, Jose-Norberto Mazónand Juan Trujillo (2013). *International Journal of Data Warehousing and Mining (pp. 46-72).*
www.irma-international.org/article/bpmn-based-design-maintenance-framework/78375

Data Mining Using Fuzzy Decision Trees: An Exposition from a Study of Public Services Strategy in the USA
Malcolm J. Beynonand Martin Kitchener (2010). *Data Mining in Public and Private Sectors: Organizational and Government Applications (pp. 47-66).*
www.irma-international.org/chapter/data-mining-using-fuzzy-decision/44282

Semi-Supervised Sentiment Classification on E-Commerce Reviews Using Tripartite Graph and Clustering
Xin Lu, Donghong Gu, Haolan Zhang, Zhengxin Song, Qianhua Cai, Hongya Zhaoand Haiming Wu (2022). *International Journal of Data Warehousing and Mining (pp. 1-20).*
www.irma-international.org/article/semi-supervised-sentiment-classification-on-e-commerce-reviews-using-tripartite-graph-and-clustering/307904