Chapter 4
REPERIO:
A Flexible Architecture for Recommendation in an Industrial Context

Frank Meyer  
Orange Labs Lannion, France

Isabelle Tellier  
LIFO Orléans, France

Damien Poirier  
LIFO Orléans, France

Françoise Fessant  
Orange Labs Lannion, France

ABSTRACT
In this chapter, the authors describe Reperio, a flexible and generic industrial recommender system able to deal with several kinds of data sources (content-based, collaborative, social network) in the same framework and to work on multi-platforms (Web service in a multi-user mode and mobile device in a mono-user mode). The item-item matrix is the keystone of the architecture for its efficiency and flexibility properties. In the first part, the authors present core functionalities and requirements of recommendation in an industrial context. In the second part, they present the architecture of the system and the main issues involved in its development. In the last part, the authors report experimental results obtained using Reperio on benchmarks extracted from the Netflix Prize with different filtering strategies. To illustrate the interest and flexibility of the architecture, they also explain how it is possible to take into account, for recommendations, external sources of information. In particular, the authors show how to exploit user generated contents posted on social networks to fill the item-item matrix. The process proposed includes a step of opinion classification.

INTRODUCTION
We are entering an era of huge catalogs and databases where one person cannot consider himself/herself to have an overview of what is available and what might interest him/her. Two types of systems were developed in parallel with the rapid inflation of the available content: search engines, and automatic recommendation systems. Search engines are serving people who know what they want and who will, more or less easily, perform a search query. Automatic recommendation systems are often used as a support system for discovery, navigation aid or as a support system for decision-
making. Aid to discovery will be done by issuing personal recommendations. The navigation aid will be done by providing a contextual help such as similar products to the product being viewed. The support for decision making will be carried out by predicting, for a user, the potential value a product may have for him/her, for instance by predicting a rating for this product.

Automatic recommendation systems exist on subjects as diverse as Web pages, movies, TV programs, books, restaurants, humorous jokes, songs, people within a social network, etc. (Rao & Talwar, 2008) identified 96 recommendation systems on various subjects, academic or industrial. The general goal of automatic recommender systems is to help users to find products (items) that should interest them, from large catalogs. Items are defined as any object that can be consumed, bought, read or viewed.

Several recommender system typologies are known in the literature: the classical typology, with collaborative filtering, content-based filtering and hybrid filtering, for instance used in Adomavicius and Tuzhilin (2005), the typology of Rao and Talwar (2008) depending on the data sources used by the systems, the typology of Su and Khoshgoftaar (2009) restricted to collaborative filtering. In the first part of this chapter, we briefly review these typologies.

However, although the literature focuses on the score prediction, industry prioritization of functionalities is very different. In this context, we introduce another possible classification of the recommender systems, based on the types of functions available, rather than on data sources. This new typology distinguishes scoring-oriented systems and those similarity-based or recommendation-oriented. We show that systems based on item-item matrix are able to cover the all functionalities an industrial recommender system needs. A system that operates in industrial context should also satisfy some requirements like robustness, scalability and reactivity, adaptation to diverse sources of data, cold-start management, etc. We describe an operational recommender system that integrates core functionalities of the recommendation and industrial requirements.

Next, we evaluate the efficiency of our industrial recommender system with an application in the field of movie recommendation. We built a benchmarks with data extracted from the Netflix challenge. We compare several filtering strategies with several kinds of sources (collaborative, content-based). As expected, the collaborative filtering approach is more efficient than the content-based one.

Finally, we also conducted experiments to exploit an original source of data: posts of users on a social network Web site concerning movies. The posts are short texts written in an SMS-like style. We show that they nevertheless express opinions that can be automatically recognized, and that these opinions can help to fill the item-item matrix on which the recommendations are based. Our experiments show that a collaborative filtering based on such a matrix is more efficient than a content-based filtering. This strategy opens interesting perspective to face the well-known cold-start problem.

**MAIN TYPOLOGIES OF RECOMMENDER SYSTEMS**

The typology currently found in the literature identifies three types of recommender systems: the collaborative filtering systems, the content-based (or thematic) filtering systems, the hybrid systems, using both collaborative and content-based techniques. This typology is based on the type of information used as input for the models (Adomavicius & Tuzhilin, 2005). Collaborative filtering uses logs of users, generally user ratings on items, sometimes with dates. Content-based filtering uses item metadata (i.e. intrinsic characteristics of the items, described by text, structural