Chapter 59
Visual Data Mining for Collaborative Filtering:
A State-of-the-Art Survey

Marenglen Biba
University of New York Tirana, Albania

Narasimha Rao Vajjhala
University of New York Tirana, Albania

Lediona Nishani
University of New York Tirana, Albania

ABSTRACT

This book chapter provides a state-of-the-art survey of visual data mining techniques used for collaborative filtering. The chapter begins with a discussion on various visual data mining techniques along with an analysis of the state-of-the-art visual data mining techniques used by researchers as well as in the industry. Collaborative filtering approaches are presented along with an analysis of the state-of-the-art collaborative filtering approaches currently in use in the industry. Visual data mining can provide benefit to existing data mining techniques by providing the users with visual exploration and interpretation of data. The users can use these visual interpretations for further data mining. This chapter dealt with state-of-the-art visual data mining technologies that are currently in use apart. The chapter also includes the key section of the discussion on the latest trends in visual data mining for collaborative filtering.

INTRODUCTION

Researchers are struggling to explore large volumes of data as the volume of data generated increased exponentially over the last few years. The traditional data mining techniques are not adequate to analyze and explore these large volumes of data as the available data is available in different dimensions and varying formats, including multimedia, geographical, and temporal data. Visual data mining techniques and approaches supplementing traditional data mining techniques can help in dealing with the large volume

DOI: 10.4018/978-1-5225-1837-2.ch059

of data. Data Mining is defined as the process of analyzing large information repositories for deriving and discovering useful patterns and identifying hidden relationships among the data (Siguenza-Guzman, Saquicela, Avila- Ordóñez, Vanderwalle & Cattysse, 2015). Data mining is used for knowledge classification and exploring consistent patterns from large volumes of collected data. Visual data mining can be quite useful and helpful in designing models for practical solutions for complex problems (Kashwan &Velu, 2012). Visual data mining has the potential to simplify and ease exploration of these large volumes of data as the user is directly involved in the data mining process (Keim, 2002).

Classification of visual data mining techniques is based on three criteria, namely, the data to be visualized, the technique of visualization, and the adopted distortion technique (Keim, 2002). Visual data mining methods can be categorized into two categories, namely, data visualization and information visualization (Kashwan &Velu, 2012). Data visualization involves the presentation of data in schematic forms, including histograms, scatter plots, and charts. Information visualization is suitable for datasets that lack standard mapping of abstract data onto the physical screen space and include visualization techniques (Keim, 2002).The data mining process includes a sequence of steps, the first of which deals with integration of raw data from different data sources, including data in data formats. Data cleaning process follows the data integration process. During data cleansing, noise, duplication and inconsistent data are removed (Siguenza-Guzman et al., 2015). The third phase is transformation into other formats that can be interpreted by various data mining tools which apply filtration and aggregation to derive summarized data. The data analyst can now derive meaningful patterns using these data mining tools. Visualization can be applied to present data to the user in a comprehensible manner.

Collaborative filtering also referred to as social information filtering, is a variant of memory-based reasoning that is suitable for application of providing personalized recommendations (Linoff & Berry, 2011). Collaborative filtering methods utilize the past ratings of users to predict or recommend new contact that the user might like (Nilashi et al., 2013). Collaborative filtering methods are often classified into two categories, namely, user-based collaborative filtering and item-based collaborative filtering. Collaborative filtering is based on the concept of similarity coupled with preferences.

Traditional recommendation systems have used collaboration filtering for making recommendations to users on the basis of how other users have rated the items. Collaborative filtering uses a three-step process for preparing recommendations for new customers (Linoff & Berry, 2011). The customer profile is first built followed by a comparison of the new customer profile with profiles of other customers using some measure of similarity. Such similar profiles are referred to as neighbor profiles in collaborative filtering. Recommendations are then made based on the predictions from the combination of customer ratings with those of the neighbor profiles (Bobadilla, Hernando, Ortega, & Gutiérrez, 2012). Collaborative filtering techniques are normally applied on large data sets consisting of different kinds of data. Visual data mining techniques, when applied for collaborative filtering could help in discovering implicit knowledge in visual forms from these large data sets. Several visualization techniques have been applied in collaborative filtering, including Multi-Dimensional Scaling (MDS), Spring Embedder, and the Navigating Exhibitions, Annotations and Resources (NEAR).