Chapter 33
Statistical Relational Learning for Collaborative Filtering a State-of-the-Art Review

Lediona Nishani
University of New York Tirana, Albania

Marenglen Biba
University of New York in Tirana, Albania

ABSTRACT

People nowadays base their behavior by making choices through word of mouth, media, public opinion, surveys, etc. One of the most prominent techniques of recommender systems is Collaborative filtering (CF), which utilizes the known preferences of several users to develop recommendation for other users. CF can introduce limitations like new-item problem, new-user problem or data sparsity, which can be mitigated by employing Statistical Relational Learning (SRLs). This review chapter presents a comprehensive scientific survey from the basic and traditional techniques to the-state-of-the-art of SRL algorithms implemented for collaborative filtering issues. Authors provide a comprehensive review of SRL for CF tasks and demonstrate strong evidence that SRL can be successfully implemented in the recommender systems domain. Finally, the chapter is concluded with a summarization of the key issues that SRLs tackle in the collaborative filtering area and suggest further open issues in order to advance in this field of research.

INTRODUCTION

With the dramatic spread of Internet, many connections and interactions have taken place in the networks. Therefore, information relying on people reactions has experienced a dramatic growth. Social Networks, which are named due to the crucial role of humans, are concentrated in the social relations that people have among them. In order to exploit these large amounts of data, it is required designing new algorithms and appropriate techniques for gathering valuable information. The field of expertise,

DOI: 10.4018/978-1-7998-0951-7.ch033

which deals with the aforementioned challenges, is Social Network Mining *(SNM).* Its major purpose is to extract information about network objects, behaviors and activities that cannot be yielded if the entities were to be examined individually.

A social network represents a social structure of entities known as nodes, which are related by one or more specific types of interdependency, such as friendship, common interests or same opinions. Social network analysis considers social relationships based on the network theory perspective; nodes represent people, whereas ties represent relations between them. Together they form a graph-based structure, which experiences significant complexity. Social networks can be categorized in four significant types due to various domains on which they operate:

1. **Citation Networks:** Deal with storing scientific papers, the relation of their subjects across their paper and corporate authors or joint paper reference
2. **SNS (Semantic Network Service) Websites**: Concern the organization of people and their related friends or professional contact for example: Facebook, LinkedIn, Instagram, twitter, etc
3. **Social Shopping Websites**: Related to e-commerce and opinion mining exploring products
4. **Social Media Websites**: Concern the suggestions about movies, music relying on user preference

Social Network Analysis has appeared to be one of the major techniques in social networks (Wasserman & Faust, 1994) intending to establish a model that can map entities and the relations between them. The probability of two entities to build a link between them may depend on the likelihood of having similar attributes. On the other hand, two entities that have developed a connection or link between them have good probability to have similar attributes.

One of the most compelling techniques of Social Network Analysis is the Statistical Relational Learning, which have emerged when traditional statistical approaches could not treat relational and complex data. Traditional approaches consider only independent and fully distributed data; therefore, they do not take into account dependent and complex relational structure of data. Furthermore, traditional Inductive Logic Programming (ILP) and relational learning do not model noisy data, hence do not perform in the presence of uncertainty. Nowadays, systems generate real world datasets, which represent multi–relational, heterogonous and semi structured data. Data mining systems deal everyday with noisy data and uncertainty. Hence, statistical relational learning approaches help to overcome these issues. Statistical relational learning can be implemented in different domains. One important task of the SRL is dealing with the recommender systems problem. Recommender systems are personalized information tools that generate recommendations, suggestions and predictions. They are employed in different areas where users might want suggestions in order to select among many various objects. For example, online portals of news provide headlines where the reader can focus. Moreover, in e-commerce perspective, online sites generally display a short list of the products in order for the user to have an overview of what the online shop provides. Subsequently, the user can select the product and explore it in more details either purchasing it.

A web page is linked with the web server, which contains a database with items available from the site. Due to the large number of items in the database, it is deemed appropriate to show the object list, which can be selected by the user. Recommender systems employ various algorithms in order to perform recommendation tasks. The algorithms aggregate items from like-minded users, thus by eliminating the item, which have been selected before and suggest the items, which have not been rated by the user. The most well-known algorithms deployed in recommender systems are collaborative filtering and content-