PRES LUNAM
Ecole Doctorale STIM
Sciences et Technologies de l'Information et Mathématiques

Spécialité : Informatique
Laboratoire : LINA
Equipe : Équipe COnnaissance et Décision

Relational Bayeasin Networks: From learning to application in Recommender Systems

Ben Ishak Mouna
Mél : mouna.ben-ishak@etu.univ-nantes.fr

Abstract: With the widespread use of Internet, recommender systems are becoming increasingly adapted to resolve the problem of information overload and to deal with large amount of online information. Several approaches and techniques have been proposed to implement recommender systems. Most of them rely on flat data representation while most real world data are stored in relational databases.

PRMs emerge as a new family of probabilistic graphical models that allow the representation of a joint probability distribution over the attributes of a relational database. Relational Bayesian networks (RBNs) are an extension of Bayesian networks in the relational context. They enable to represent uncertainty over objects and relations while using the entire rich structure of relational databases.

In this paper, we describe our new RBN-based recommender system architecture that applies probabilistic inference over objects to provide personalized recommendations, we give an overview of our ongoing work which consists on proposing a new RBN structure learning algorithm, and we describe our development progress.

Keywords: *Relational Bayesian Network, Relational models structure learning, Recommender System.*
Collaborations : Cosupervision with LARODEC, ISG Tunis, Tunisia

1 Introduction

Recommender systems [1] are being increasingly adopted in a wide range of applications, mainly in e-commerce applications as they help increase online sales and improve customer loyalty. Several approaches and techniques have been proposed to implement recommender systems. One line of research aims to cast the recommendation task into a classification problem and use standard data mining techniques to perform recommendation. However, almost all used techniques rely on propositional data representation.

Relational databases representation dominates computer industry mainly for storing and retrieving data. Recently, there has been growing interest in extracting interesting statistical patterns from relational models which are the most common representation of structured data. Namely, Probabilistic Relational Models (PRMs) emerge as a new family of probabilistic graphical models to represent uncertainty over the properties of an entity, capturing its probabilistic dependence both on other properties of that entity and on properties of related entities.

Our purpose is to make contributions in both areas: First, we proposed a new hybrid recommendation approach, based on relational Bayesian networks. Our architecture rely on the entire rich structure of relational databases instead of the flat data representation. Second, we are currently developing a new algorithm for learning the structure of RBNs.

The rest of the paper is as follows. Sections 2 and 3 introduce our working axes, respectively recommender systems and relational Bayesian networks, and highlight the motivation for the envisaged contributions. Section 4 represents our RBN-based recommender system and finally Section 5 concludes.

2 Recommender systems

Recommender systems [1] emerged in the mid-1990s as a new research area whose interest has increased recently with the intension of reducing part of the information overload problem produced on the Net. They are invoked in many Internet sites such as Amazon, YouTube, Yahoo, Netflix, etc. In 2009, Netflix awarded a million dollar prize to the team that first succeeded in improving substantially the accuracy of predictions of its recommender system[1]. The ultimate goal of a recommender system is to deliver a list of personalized recommended items to a particular user within a specific domain.

[1]http://www.netflixprize.com/

Several recommender systems have been developed. Nonetheless, collaborative filtering and content-based approaches stay the most familiar and mature ones. The former attempts to identify groups of users with similar tests as the active user and recommends items that they have liked. The latter learns to recommend items that are similar to those the user has liked in the past. Some research established trade-offs between these two approaches in order to provide hybrid systems that overcome the shortcomings of each [2].

Recommender systems development involves various disciplines such as Human Computer Interaction, Information Retrieval, Marketing, etc. However, most of these systems are typically centered around the use of various machine learning algorithms to predict user evaluations for items, or for learning how to correctly rank items for a user. In this context, classification and clustering techniques are quite applied and Bayesian networks are among the most important classification techniques used for recommendation. Often these algorithms rely on propositional data representation.

Relational data base representation is a natural data structure representation to store and easily and efficiently handle large data sets. So, we propose a new hybrid recommendation approach, based on relational Bayesian networks. Our architecture rely on the entire rich structure of relational databases instead of the flat data representation. Before describing our approach, we give a brief representation of RBNs.

3 Relational Bayesian Networks

Probabilistic relational models are classified into three main groups depending on their graphical representation: Relational Bayesian Networks (RBNs) representing directed acyclic graphs [3, 4], Relational Markov Networks (RMNs) representing undirected graphs [5] and Relational Dependency Networks representing bi-directed graphs [6].

In our work we focus on RBNs which are an extension of Bayesian networks (BNs) [7] in the context of relational data, where the probability model specification concerns classes of objects rather than simple attributes.

3.1 Model Definition

A RBN Π for a relational schema $\mathcal{R}$ (i.e., set of entities and relations) is defined through a qualitative dependency structure $\mathcal{S}$ and a set of parameters associated with it $\theta_{\mathcal{S}}$. The relational schema $\mathcal{R}$ describes a set of classes $\mathcal{X} = \{X_1, \ldots, X_1\}$, each of which has a set of descriptive attributes denoted by $\mathcal{A}(X)$, which take on a range of values $\mathcal{V}(X.A)$ and a set of reference slots denoted by $\mathcal{R}(X) = \{\rho_1 \ldots \rho_k\}$. Each $X.\rho$ has X as domain type and Y as a range type, where $Y \in \mathcal{X}$. A sequence of slots $\rho_1 \ldots \rho_k$, where $\forall i, Range[\rho_i] = Dom[\rho_{i+1}]$ defines a slot chain K. The notion of aggregation is also adopted from the database theory: An aggregate γ takes a multiset of values of some ground type, and returns a summary of it.

A skeleton structure σ_r of a relational schema specifies a set of objects and relations that hold between them in a specific domain, without specifying values of probabilistic attributes. Thus, given a relational skeleton σ_r, the RBN Π defines a distribution over the possible worlds consistent with σ_r.

On other words, A RBN Π together with a relational skeleton σ_r define an instance dependency structure $\mathcal{I}$ known as the ground Bayesian network $GBN =< \mathcal{G}_{\sigma_r}, \theta_{\mathcal{G}_{\sigma_r}} >$.

3.2 Relational structure learning

To perform probabilistic inference, the RBN structure have to be already defined, if not, it have to be learned from data. Learning the structure of a RBN consists on providing a qualitative dependency structure $\mathcal{S}$ which fits the best possible way to the observed data. Proposed methods are inspired from classical BN learning techniques from propositional data and are adapted to relational data. [8] proposed a score-based approach, while [9] proposed a constraint-based one.

In the context of standard Bayesian networks, Tsamardinos et al. [10] proposed an hybrid method that combines constraint-based and score-based approaches and their experimental results show that hybrid algorithms are more effective than algorithms based only on a single technique. By analogy, we attempt to propose an hybrid approach to learn RBN from relational data that we refer to as Relational Max-Min-Hill-Climbing as it represents an extension of the structure learning algorithm proposed by [10] in the context of relational data.

4 Relational recommender system

Only few works used RBNs to model recommender systems. [11] describe how to apply RBNs in the context of collaborative filtering, and emphasize their ability to deal with much more relational information available than the simple user-item relationships. [12] use RBN with class hierarchies [4] to the movie recommendation task and show that their approach achieves state-of-the-art results. [13] use RBNs in the context of collaborative filtering in order to improve recommendation quality for low grade users. [14] treat the recommendation problem as a special type of the relational learning problem. The idea is to estimate a RBN model and to specify a set of relevant attributes to perform recommendation. Most of these methods tackle the collaborative filtering issue. In what follows, we will describe a new RBN-based recommendation approach. Our approach uses RBNs to model the recommender system domain. Then, resorts to probabilistic inference to provide recommendation.

In each recommender system, a set of data about users as well as items is available: A user can rate many items. Mutually, an item can be chosen by many users. Besides, to each user we can store a set of his demographic information such as age, gender, location, etc. On the other hand, a set of items characteristics is also provided.

In our work, we suppose that the recommendation domain data is stored in a relational data base, and a RBN over its relational schema (i.e., the class dependency graph and the CPD of each variable) is constructed either by an expert or by a RBN learning algorithm.

The RBN qualitative dependency graph $\mathcal{S}$ represents two types of links:

- *Intra-class links:* connecting attributes of the same class. This type of connections allows to find dependencies among features of the same class.
- *Inter-class links:* connecting attributes of different classes. This type of connections allows to find dependencies among features of different classes.

Our RBN-based representation allows to benefit from different recommendation approaches at once. So the system does not rely exclusively on users or items similarities. Rather, different features are thrown together into a single recommendation process while optimizing the selection of the most correlated and most relevant attributes for the recommendation through the graph dependency structure. We propose a feature combination hybrid method [2], based on three steps, which are recapitulated in Fig. 1:

1. **RBN modification:** From an initial relational schema and a RBN Π associated with this schema, we construct a new RBN Π' by dividing the class associated to the ratings into two classes, namely, Sound-Votes and Forecast-Votes.
2. **RBN instantiation:** We create a relational skeleton σ_{u_a} to each active user u_a based on a set of rules. σ_{u_a} together with the set of conditional probability distributions of Π' form the ground Bayesian network of u_a.
3. **Performing recommendation:** Using an inference process we compute $P(Rating.rating|Pa(Rating.rating))$ for the Forecast-Votes objects, relevant ones are those having the highest probability value.

5 Conclusion

As a first proposal, we have emphasized the relational nature of recommender systems by the use of RBNs and designed the recommendation task using a RBN. Then, we resorted to probabilistic inference to predict user's preferences. Our approach allows the integration of relational data into the recommendation process. So, features related to users and items are exploited to provide useful information during recommendation.

We worked under the assumption that the RBN model is given as input, which is not always obvious to obtain. In our ongoing work, we work on the development of a new relational structure learning algorithm. We are implementing a C++ toolbox allowing, definition, learning, inference and visualisation of RBN using the Boost library[2] and APIs provided by the ProBT platform[3]. The learning module is under construction.

[2]http://www.boost.org
[3]http://emotion.inrialpes.fr/BP

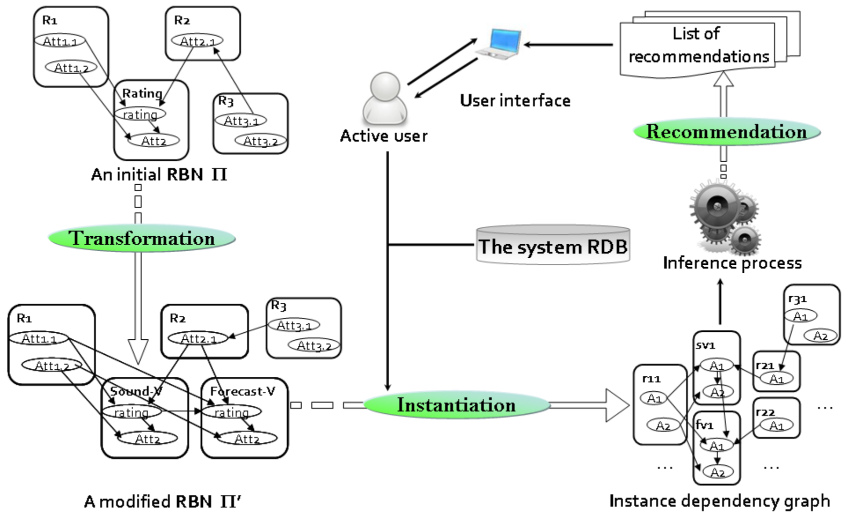

Figure 1: The overall architecture of the RBN-based recommender system

References

[1] F. Ricci, L. Rokach, B. Shapira, and P. Kanto. *Recommender Systems Handbook*. New York, USA, Springer, 2010.

[2] R. Burke. Hybrid recommender systems: Survey and experiments. *User Modeling and User-Adapted Interaction*, 124:331–370, 2002.

[3] A. J. Pfeffer. *Probabilistic Reasoning for Complex Systems*. PhD thesis, Stanford University, 2000.

[4] L. Getoor. *Learning statistical models from relational data*. PhD thesis, Stanford University, 2002.

[5] B. Taskar, P. Abbeel, and D. Koller. Discriminative probabilistic models for relational data. In *Proceedings of the 18th UAI*, pages 485–492, 2002.

[6] J. Neville and D. Jensen. Relational dependency networks. *Journal of Machine Learning Research*, 8:653–692, 2007.

[7] J. Pearl. *Probabilistic reasoning in intelligent systems*. San Franciscos, Morgan Kaufmann, 1988.

[8] N. Friedman, L. Getoor, D. Koller, and A. Pfeffer. Learning probabilistic relational models. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 1300–1309, 1999.

[9] M. Maier, B. Taylor, H. Oktay, and D. Jensen. Learning causal models of relational domains. In *Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence*, 2010.

[10] I. Tsamardinos, L. E. Brown, and C. F. Aliferis. The max-min hill-climbing bayesian network structure learning algorithm. *Machine Learning*, 65:31–78, 2006.

[11] L. Getoor and M. Sahami. Using probabilistic relational models for collaborative filtering. In *Working Notes of the KDD Workshop on Web Usage Analysis and User Profiling*, 1999.

[12] J. Newton and R. Greiner. Hierarchical probabilistic relational models for collaborative filtering. In *Proceedings of the 21st International Conference on Machine Learning, Workshop on Statistical Relational Learning*, 2004.

[13] Y. Gao, H. Hong, J. Liu, and D. Liu. A recommendation algorithm combining user grade-based collaborative filtering and probabilistic relational models. In *Proceedings of the Fourth International Conference on Fuzzy Systems and Knowledge Discovery*, 2007.

[14] Z. Huang, D. D. Zeng, and H. Chen. A unified recommendation framework based on probabilistic relational models. In *Proceedings of the fourteenth Annual Workshop on Information Technologies and Systems*, 2004.