A Unified Recommendation Framework Based on Probabilistic Relational Models

Zan Huang, Daniel D. Zeng, Hsinchun Chen
Department of Management Information Systems, The University of Arizona
{zhuang, zeng, hchen}@eller.arizona.edu

Abstract
Recommender systems are being increasingly adopted in various e-commerce applications. A wide range of recommendation approaches have been developed to analyze past consumer-product interactions, consumer attributes, and product attributes to predict future sales. In this paper we propose a unified recommendation framework based on *probabilistic relational models* (PRMs). This framework includes most of existing recommendation approaches, such as collaborative filtering, content-based, demographic filtering, and hybrid approaches, as special cases. Recently developed in the machine learning community, PRMs aim to study the relational patterns within a database containing multiple interlinked data tables using a statistical model that describes probabilistic dependencies between attributes in the domain. We extended the original PRMs in order to capture relational data patterns that are important for recommendation. We also specialized the algorithm for learning PRMs in dependency model construction and parameter estimation to exploit the special characteristics of the recommendation problem. Through an experimental study, we demonstrate that the proposed framework not only conceptually unifies existing recommendation approaches but also allows the exploitation of a wider range of relational data patterns in an integrated manner, leading to improved recommendation performance.

1. Introduction

Recommender systems are being widely used in many application settings to suggest products, services, and information items to potential consumers. A wide range of companies such as *Amazon.com*, *Half.com*, *CDNOW*, *J.C. Penney*, and *Procter & Gamble* have successfully deployed recommendation technologies to increase Web and catalog sales and improve customer loyalty [14]. A variety of recommendation approaches have been developed in the Artificial Intelligence and Information Retrieval communities [1, 10, 13]. Most of theses approaches take as input three types of data: product attributes, consumer attributes, and interactions between consumers and products (such as purchases and ratings). As output, they predict future or unobserved interactions as recommendations. Depending on the input data these approaches can be roughly categorized into content-based (using product attributes and the interaction data), collaborative filtering (using the interaction data only), demographic filtering (using consumer attributes and the interaction data), and hybrid approaches (using multiple types of input data) [7, 10, 13].

Conceptually, the recommendation problem is concerned with the relationships between consumers and products. As such, it can be viewed as a special case of the *relational learning* problem [3]. Recent years have seen significant interest and development in the area of relational learning, which focuses on identifying relational patterns within a database containing multiple interlinked data tables. Applying the relational learning framework, one can argue that a recommendation model takes a (portion of the) database containing multiple related tables regarding consumers, products, and their interactions as input to predict unobserved entries in the consumer-product interaction table.

The main objective of this paper is to establish the connection between the recommendation problem and the relational learning framework through the application of a recently developed statistical relational learning method called *probabilistic relational models* (PRMs) in the recommendation context. We extend the original PRMs to meet the unique computational challenges of the recommendation task. We show that existing recommendation approaches can be conceptualized as special cases under this framework. We also demonstrate the improved recommendation performance achieved by the unified framework using a real-world e-commerce dataset.

2. Relational Learning and Probabilistic Relational Models

Relational learning or *multirelational learning* [3] extends standard data mining that learns from attributes of independent entities stored in a single database table to extract patterns from multiple related tables. The assumption that the data objects are independent from each other is dropped in relational learning. In fact, linkages between data objects are of central interest in relational learning. Examples of relational learning applications include link prediction, link-based clustering, social network modeling, and object identification.

Probabilistic relational models (PRMs) are the main formal approach that has been developed for relational learning [8, 12]. A PRM is defined for a particular database, or formally a *relational schema R*. A relational schema R describes a set of classes (tables in the database) $\boldsymbol{X}$. Each $X \in \boldsymbol{X}$ is associated with a set of *descriptive attributes* (standard table attributes) $\boldsymbol{A}(X)$ and a set of *reference slots* (foreign keys) $\boldsymbol{R}(X)$. We denote the attribute A of class X as $X.A$ and the reference slot ρ of X as $X.\rho$, where ρ denotes a function from Domain[ρ] = X to Range[ρ] = Y. PRMs with *existence uncertainty* [5] are able to model the existence of certain records in the data tables. Under this extension, a class X of interest can be modeled as an *undetermined* class by introducing a special *existence* attribute $X.E$ whose values are from $V(E) = \{true, false\}$, with *true* indicating the particular object of class X exists and *false* indicating nonexistence. For each reference slot ρ we define an inverse reference slot ρ^{-1}, mapping from Range[ρ] = Y to Domain[ρ] = X. A slot chain τ is defined as $\tau = \rho_1,\ldots,\rho_k$, for all i, Range[ρ_i] = Domain[ρ_{i+1}]. Through slot chains dependencies between the attributes of related data objects can be explored.

A PRM is an extension of *Bayesian networks* for describing probability distributions over a database. A PRM Π contains a qualitative component S, an acyclic graph that describes the statistical dependency structure of descriptive attributes linked through slot chains, and a quantitative component Θ_S that represents the set of parameters characterizing the conditional probability distributions. Formally, a PRM $\Pi = \langle S, \Theta_S \rangle$ for a *relational schema* $R = \langle \boldsymbol{X}, \boldsymbol{A} \rangle$ defines for each class $X \in \boldsymbol{X}$ and each descriptive attribute $A \in \boldsymbol{A}(X)$, a set of parents $Pa(X.A)$, and a conditional probability distribution that represents $P(X.A \mid Pa(X.A))$ [5]. A complete *instantiation I* for a PRM is defined as the set of objects in each class X and the values for each attribute and each reference slot of each object. With a complete instantiation a PRM can be learned by finding a PRM Π that best matches I. Similar to Bayesian network learning, a statistically motivated scoring function is used to evaluate each model with respect to the training data. A commonly used Bayesian scoring metric is given by $\log P(S \mid I) = \log P(I \mid S) + \log P(S) + C$, where $P(I \mid S)$ is the marginal likelihood $P(I \mid S) = \int P(I \mid S, \Theta_S) P(\Theta_S \mid S) d\Theta_S$. Standard hill-climbing greedy search algorithms can be employed to search for the optimal structural model S. With the optimal dependency structure, standard *maximum likelihood* parameter estimation can be performed to complete the model specification. Details on PRM learning can be found in [4, 5].

3. PRM-based Recommendation

The recommendation problem is an ideal application for relational learning as the linkages between consumers and products are the modeling focus. In fact one most successful recommendation approach, collaborative filtering, makes recommendation predictions only based on these linkages. In this section, we establish the connection between recommendation problem and relational learning. Using a book sales database as an example, Section 3.1 describes recommendation task from a relational learning perspective. Section 3.2 introduces an extension to the original PRM, motivated to evaluate similarities between sets required by the recommendation model. Section 3.3 presents a simplified parameter estimation procedure exploiting the characteristics of recommendation tasks. We demonstrate in Section 3.4 that existing recommendation approaches can be viewed as special cases of our unified framework.

3.1 Model Description

Figure 1 illustrates an example book sales database. Customer, Book, and Word are entity classes while Order and Occurrence are relationship classes. Customer and Book have 7 and 3 descriptive attributes, respectively, excluding their identifiers. Order contains two references slots linking to Customer and Book while Occurrence contains book and word reference slots describing the occurrence of keywords in book content descriptions such as title and introduction. For recommendations, we model Order as an undetermined class by introducing a special descriptive attribute exist to indicate the existence of a sales transaction. In Figure 1, we only present existing records (with 1 assigned to the exist attribute) while for all other customer-book pairs the value 0 is implicitly assigned indicating the absence of the sale record.

Customer							
customer	city	birthYear	education	vocation	sex	married	child
c1	taipei	1977	college	financial	f	yes	1
c2	kaohsuing	1968	high school	construction	m	no	0
c3	taipei	1982	college	student	m	no	0

Order		
customer	book	exist
c1	b1	1
c1	b2	1
c2	b3	1
c3	b4	1

Book			
book	publisher	translated	price
b1	p1	yes	130
b2	p1	yes	230
b3	p2	no	100
b4	p3	no	500

Occurrence	
book	word
b1	w2
b1	w3
b2	w1
b3	w4
b3	w5
b4	w4
b4	w6

Word
word
w1
w2
w3
w4
w5
w6

Figure 1. An example book sales database

Using the PRM notation introduced in Section 2, the dependency structure S for a PRM defines the parents $Pa(X.A)$ for each attribute $X.A$. In our context, since we are only concerned with Order.exist, we only need to derive a partial dependency structure for Order.exist. Instead of searching for the complete model describing all probabilistic dependencies we focus on identifying attributes within the *Markov blanket* of Order.exist and search for an optimal model describing these variables and Order.exist. A Markov blanket refers to the parents, children, and other parents of the children of a node V in a Bayesian network model; it shields V from being affected by any node outside the blanket [2]. Potential attributes to be included into the Markov blanket of Order.exist can be derived from reference slots or slot chains. For example, [Order.customer].education could be a potential parent attribute representing the education level of the target customer. Long slot chains with inverse reference slots can bring in more complex attributes. For example, [Order.customer].[Order.customer]$^{-1}$.[Order.book].price represents the prices of the set of books bought by the target customer. If any of the reference slots in the chain involves a one-to-many mapping, such as [Order.customer]$^{-1}$ (indicating the function from a customer to his/her involved orders) in this example, the derived attribute will be a multi-valued attribute. For these attributes, aggregation operators such as *maximum, minimum, mode, average,* and *carnality* can be applied and aggregated single-valued attributes are then included into the dependency structure S.

3.2 Multi-set Operations

A PRM allows attributes to be derived separately from individual slot chains. For example, [Order.customer].[Order.customer]$^{-1}$.[Order.book].[Order.book]$^{-1}$.[Order.customer] represents the set of customers who bought at least one common book bought by the target customer (the customer neighbors) while [Order.book].[Order.book]$^{-1}$.[Order.customer] represents the set of customers who bought the target book. These two multi-valued attributes, with aggregation operations, could provide certain information regarding the likelihood for a transaction involving the target consumer and target book to occur. However, the original PRMs do not model information that can only be derived *jointly* from multiple attributes, which can play a critical role in recommendation. For example, the set similarity of the above two attributes (may be derived through the cardinalities of the intersection and union of the two attributes) describes the overlap of the target consumer's neighbors and the customers who bought the target book. Such information is essential for making user-based collaborative filtering recommendations.

To derive information jointly from multiple attributes, we propose to extend PRMs by introducing *multi-set operators*. A multi-set operator ϕ_k on k multi-valued attributes A_1, …, A_k that share the same domain $V(A_1)$ denotes a function from $V(A_1)^k$ to $V(A_1)$. Such multi-set operators include simple set operators such as *intersection* and *union* and more complex aggregation operators modeling value distributions [11]. By applying an aggregation operator after a multi-set operator, we can derive attributes from multiple multi-valued attributes and include them into the probabilistic dependency specification of a PRM. Our current study focuses on using binary multi-set operators that involve two attributes.

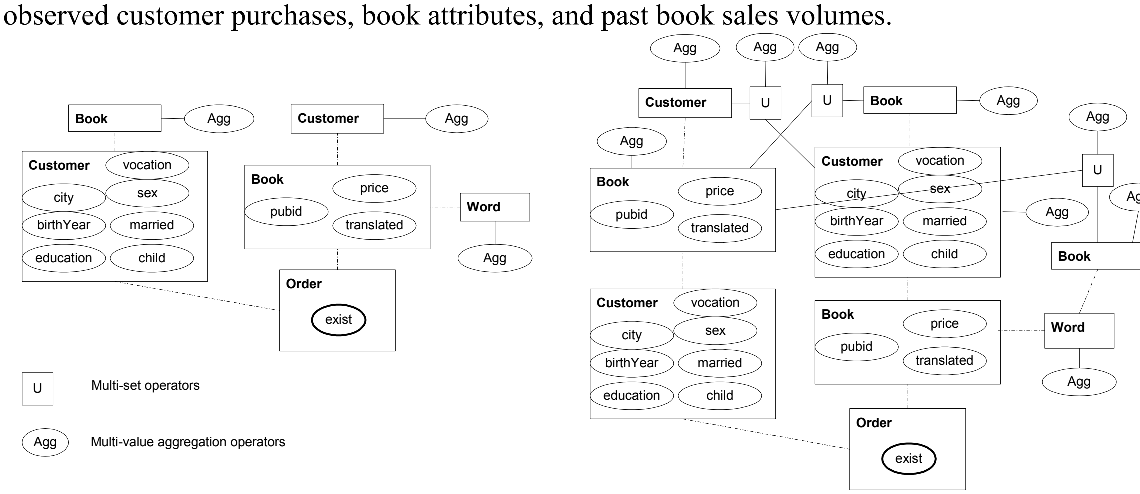
3.3 Learning Process

An important challenge for PRM learning is that there are an infinite number of potential attributes that could be derived through slot chains and the newly introduced multi-set operators. The standard approach to address this issue is an iterative expanding heuristic structure search algorithm [4]. Under this approach, the length of the slot chains is constrained while searching for the optimal PRM. The allowable slot chain length controls the complexity of the model.

We applied standard search-and-scoring procedure for Bayesian network learning to search the optimal partial dependency structure involving Order.exist as discussed in Section 3.1. Once the optimal local PRM dependency structure for a particular slot chain length is determined, standard parameter estimation procedures can be applied to derive predictive models of the value of Order.exist. In our current study, a Naïve Bayesian algorithm for binary prediction [2] was applied to estimate the purchase probability *P*(Order.exist=1 | relevant attributes of Order.exist) for unobserved customer-book pairs. Various types of recommendations can then be generated based on such purchase probability estimates.

3.4 Recommendation Models under PRM

We now examine existing recommendation approaches in light of our unified PRM framework. For illustration purposes, we use the same book sales database example. In Figure 2, attributes in circles, including single-valued and aggregated multi-valued attributes, are potential relevant attributes of Order.exist in the dependency model. Model (a) includes attributes derived through slot chains with the maximum length of 3 (e.g., city of the target customer [Order.customer].city and number of customers who bought the target book *cardinality*{[Order.book].[Order.book]$^{-1}$.[Order.customer]}). Such a model corresponds to typical purchase prediction models that involve customer demographic attributes, number of observed customer purchases, book attributes, and past book sales volumes.



(a) Maximum length of slot chains: 3 (b) Maximum length of slot chains: 5

Figure 2. PRM recommendation model with maximum slot chain lengths of 3 and 5

Model (b) allows for slot chains of the maximum length of 5. As the model gets more complicated, many interesting attributes that correspond to existing recommendation approaches appear. As explained in Section 3.2, *cardinality*{*intersection*{[Order.customer].[Order.customer]$^{-1}$.[Order.book].[Order.book]$^{-1}$.[Order.customer], [Order.book].[Order.book]$^{-1}$.[Order.customer]}} represents the number of the target customer's neighbor customers who have bought the target book, which provides essential information for the standard user-based collaborative filtering algorithm. Similarly, *cardinality*{*intersection* {[Order.customer].[Order.customer]$^{-1}$.[Order.book], [Order.book].[Occurrence.book]$^{-1}$.[Occurrence.word].[Occurrence.word]$^{-1}$.[Occurrence.book]}} represents the number of books bought by the target customer that contain words appearing in the target book, which provides essential information for content-based recommendation approaches. The PRM estimated based on model (b) could potentially be a "hybrid" recommendation based on multiple algorithmic implementations of several recommendation approaches. As the maximum slot chain length increases, the model becomes more complicated by introducing indirect customer/book neighbors and their associated attributes, which in principle correspond to graph-based recommendation algorithms that account for transitive associations among customers and books leading to better recommendation performances (e.g., [6]). However, a larger maximum slot chain length also leads to a dramatically larger search space, making the model estimation and prediction process much more computationally intensive.

4. Experimental Study

We used a book sales dataset from a major Chinese online bookstore for our experimental study. This dataset covers 5 years of transactions of a sample of 2,000 customers, involving 9,695 books and 18,771 transactions. We derived from the raw data a database with a schema corresponded to the example in Figure 1. For the word occurrence table we only included indexed phrases in book titles and keywords.

Due to the space limit we only report a subset of our experimental results in this paper. We estimated PRMs with maximum slot chain lengths of 3 and 5 corresponding to our discussions in Section 3.4. Table 1 presents the relevant attributes of Order.exist under models (a) and (b). Book attributes did not contribute to explain the occurrences of sales records in model (a) while a wide variety of different types of derived features were included into the optimal parent set for model (b).

Model	Maximum length of slot chains	Relevant attributes
(a)	3	c's number of past purchases, city, vocation, birthYear, education, sex, child
(b)	5	number of c's neighbors that have bought b, number of b's neighbors that c has bought, number of books bought by c that contain words appear in b, number of books bought by c that are published by the publiser of b, number of books bought by c that are at the same price level as b, number of customers who bought b and have the same vocation as c, number of customers who bought b and have the same education level as c, number of customers who bought b and were born in the same decade as c was born.

*We denote the target customer as c and target book as b

Table 1. Relevant attributes to Order.exist of models (a) and (b)

To provide recommendation performance evaluation consistent with the literature, we generated top-N recommendations based on the purchase probability estimates obtained from our model. For each customer we recommended top 10 books that he/she had not purchased previously ranked by the estimated purchase probabilities. In addition to this typical recommendation task, other recommendation tasks can be also supported, such as finding the most likely customers for a particular book and finding the orders that are most likely to occur.

We report in Table 2 top-N recommendation performance for using subsets of the attributes included in model (b) that correspond to existing recommendation approaches and the performance using all attributes in model (b). We employed standard top-N recommendation quality measures including precision, recall, F measure, and rank score [1, 10] to evaluate the accuracy, coverage, and ranking quality by matching the recommendation lists with withheld 30% later actual purchase records. The experimental results showed that the complete optimal model achieved higher recommendation quality measures than all models corresponding to existing recommendation approaches. We also observed that the model based on demographic attributes had the second best performance for our dataset. [1]

Relevant attributes	Corresponding recommendation approach	Precision	Recall	F	Rank Score
number of c's neighbors that have bought b	user-based CF	0.0197	0.0491	0.0250	3.9450
number of books bought by c that contain words appear in b	content-based	0.0195	0.0381	0.0228	2.6418
number of customers who bought b and have the same vocation as c, number of customers who bought b and have the same education level as c, number of customers who bought b and were born in the same decade as c	demographic filtering	0.0265	0.0687	0.0343	6.4160
all relevant attributes shown in Table 1	(complete model)	0.0308	0.0817	0.0399	6.4498

Table 2. Recommendation performance of sub models and the complete model under model (b)

5. Conclusions and Future Directions

In this paper we have presented a unified recommendation framework based on probabilistic relational models treating the recommendation problem as a special type of relational learning problem. We

[1] For fast prototyping we implemented the PRM learning algorithms using MS SQL Server stored procedures, which are quite inefficient in both computing time and memory requirement. For our testing dataset, this implementation took more than 6 hours to complete the feature construction and selection processes while the parameter estimation and prediction generation processes took less than 5 minutes. We expect significant reduction in computing time and space requirement with more efficient programming environments.

extended PRMs by introducing multi-set operators to explore the important relational data patterns relevant to recommendation. We also specialized the dependency structure searching and parameter estimation procedures to exploit the characteristics of recommendation problem. The proposed PRM-based recommendation framework allows the integration of various existing recommendation approaches as well as exploitation of a wider range of relational data patterns. Our experimental results using data provided by an online bookstore showed that such a unified framework resulted in improved recommendation performance.

We are in the process of extending our work in the following directions: (a) application of complex multi-set aggregation operations that provide richer input information for learning; (b) optimization of the algorithm implementation to improve space and time efficiencies; (c) a complete comparison with a wide range of existing recommendation approaches. Recommendation is only one special application of PRM-based relational learning. We can also develop other specializations of PRMs for a wide range of personalization, marketing, and managerial applications based on consumer behavior modeling [9].

6. Acknowledgement

This research is partly supported by an NSF Digital Library Initiative-II grant, "High-performance Digital Library Systems: From Information Retrieval to Knowledge Management," IIS-9817473, April 1999 – March 2002, and by an NSF Information Technology Research grant, "Developing a Collaborative Information and Knowledge Management Infrastructure," IIS-0114011, September 2001 – August 2005.

7. Reference

[1] J.S. Breese, D. Heckerman and C. Kadie, Empirical Analysis of Predictive Algorithms for Collaborative Filtering, in *Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence*, 43-52, (1998).

[2] J. Cheng and R. Greiner, Comparing Bayesian Network Classifiers, in *Proceedings of the 15th Conference on Uncertainty in Artificial Intelligence*, 101-107, (1999).

[3] S. Dzeroski and N. Lavrac, Relational Data Mining (Springer-Varlag, Berlin, 2001).

[4] N. Friedman, L. Getoor, D. Koller and A. Pfeffer, Learning Probabilistic Relational Models, in *Proceedings of the 16th International Joint Conference on Artificial Intelligence*, 1300-1307, (1999).

[5] L. Getoor, N. Friedman, D. Koller and B. Taskar, Learning Probabilistic Models of Link Structure, *Journal of Machine Learning Research*, 3, 679-707, (2002).

[6] Z. Huang, H. Chen and D. Zeng, Applying Associative Retrieval Techniques to Alleviate the Sparsity Problem in Collaborative Filtering, *ACM Transactions on Information Systems*, 22(1), 116-142, (2004).

[7] Z. Huang, W. Chung and H. Chen, A Graph Model for E-Commerce Recommender Systems, *Journal of the American Society for Information Science and Technology*, 55(3), 259-274, (2004).

[8] D. Koller and A. Pfeffer, Probabilistic Frame-Based Systems, in *Proceedings of the 15th Conference of the American Association for Artificial Intelligence*, 580-587, (1998).

[9] B.P.S. Murthi and S. Sarkar, The Role of the Management Sciences in Research on Personalization, *Management Science*, 49(10), 1344-1362, (2003).

[10] M. Pazzani, A Framework for Collaborative, Content-Based and Demographic Filtering, *Artificial Intelligence Review*, 13(5), 393-408, (1999).

[11] C. Perlich and F. Provost, Aggregation-Based Feature Invention for Relational Learning, in *Proceedings of the 9th SIGKDD International Conference on Knowledge Discovery and Data Mining*, (2003).

[12] D. Poole, Probabilistic Horn Abduction and Bayesian Networks, *Artificial Intelligence*, 64, 81-129, (1993).

[13] P. Resnick and H. Varian, Recommender Systems, *Communications of the ACM*, 40(3), 56-58, (1997).

[14] J. Schafer, J. Konstan and J. Riedl, E-Commerce Recommendation Applications, *Data Mining and Knowledge Discovery*, 5(1-2), 115-153, (2001).