generated from the topic-specific latent vectors in each topic. Our model characterizes the core ideas in Eq. 1 in a Bayesian probabilistic way: each topic can be considered as a cluster. Using a multi-topic latent representation, our model is more powerful to reflect the complex characteristics for users and items in rating prediction. Another important merit of using topics is that our approach has a better model interpretability. Since topic models are effective to discover coherent topical semantics [Blei *et al.*, 2003], the derived topics in our model also group the items that are highly correlated together. In this way, a topic will be more coherent than what have been obtained in previous studies [Mackey *et al.*, 2011; Lee *et al.*, 2013; 2014; Chen *et al.*, 2015]. With topics as contextual information, we can analyze how the rating preference of a user varies in different topical contexts.

Our work presents a Bayesian formulation of local matrix factorization for the first time, which elegantly combines topic models with probabilistic matrix factorization models. By using topics as clusters, our approach has a better model interpretability. Extensive experiments on large real-world datasets demonstrate the effectiveness of the proposed model compared with several competitive baselines.

2 Related Work

In this section, we review the related work.

Matrix Factorization. MF [Paterek, 2007; Mnih and Salakhutdinov, 2007; Koren *et al.*, 2009] is an important kind of model-based collaborative-filtering methods. MF constructs low-rank approximation by projecting users and items into a latent low-dimensional space. Further, Probabilistic Matrix Factorization (PMF) has been proposed by using the Gaussian distribution to model observed ratings with zero-mean spherical Gaussian priors. In essence, PMF can be considered as a probabilistic realization of Regularized Singular Value Decomposition. Further, Salakhutdinov and Mnih [Salakhutdinov and Mnih, 2008] presented a full Bayesian formulation of PMF. As the extensions of MF, biased MF and SVD++ have been proposed in [Koren *et al.*, 2009]. Biased MF incorporates both user bias and item bias, while SVD++ uses implicit feedback to improve user preference modeling.

Local Matrix Factorization. Recently, several studies focus on using the ensemble of submatrices for better low-rank approximation, including DFC [Mackey *et al.*, 2011], LLORMA [Lee *et al.*, 2013; 2014], ACCAMS [Beutel *et al.*, 2015] and WEMAREC [Chen *et al.*, 2015]. These methods partition the original matrix into several smaller submatrices, and a local MF is applied to each submatrix individually. The final predictions are obtained using the ensemble of multiple local MFs. Typically, clustering-based techniques with heuristic adaptations are used for submatrix generation. We give a brief review of these studies. Mackey et al. [Mackey *et al.*, 2011] introduces a Divide-Factor-Combine (DFC) framework, in which the expensive task of matrix factorization is randomly divided into smaller subproblems. LLORMA [Lee *et al.*, 2013; 2014] uses a non-parametric kernel smoothing method to search nearest neighbors; WEMAREC [Chen *et al.*, 2015] employs Bregman co-clustering [Dhillon *et al.*, 2003] techniques to partition the original matrix; ACCAMS adopts an additive co-clustering approach [Shan and Banerjee, 2008] to derive sub-matrices and predict the ratings using a Gaussian distribution. Our work is highly built on the above studies, however, we propose to use probabilistic topic models to create "soft" clusters, further develop a full Bayesian model by integrating topic models with probabilistic MF.

Matrix Factorization with Topic Models. In the literature, researchers have made several attempts to combine topic models with MF, including CTM [Wang and Blei, 2011], HFT [McAuley and Leskovec, 2013], and ETF [Zhang *et al.*, 2014]. However, these methods mainly aim to incorporate ratings into topic models, and focus on combining the merits from both kinds of models. Typically, a single MF component is used, which is not suitable for local MF.

3 Bayesian Probabilistic Multi-Topic Matrix Factorization

In this section, we present our model BPMTMF for rating prediction. A glossary of notations used in the paper are listed in Table 1. In what follows, we denote matrices by bold capital letters. Superscripts, such as in $\mathbf{P}^{(k)}$, denote different topics' matrices for different superscripts; Subscripts on matrices denote the indices of data. For example, $\mathbf{R}_{um}$ denotes the entry in the u-th row and m-th column of the k-th topic matrix.

Table 1: Notations used in the paper.

Symbols	Descriptions
N, M	number of rows (users) and columns (items)
$\mathbf{R}$	data matrix ($\in \mathbb{R}^{N\times M}$) (with missing values)
K	the number ($\ll \min(N, M)$) of topics (or *topic number* for simplification)
D	the number ($\ll \min(N, M)$) of dimensions for latent vectors
$i = \langle u, m\rangle$	the i-th observation in $\mathbf{R}$
$\mathbf{P}_u^{(k)}$	the topic-specific latent vector ($\in \mathbb{R}^D$) for the u-th user *w.r.t.* the k-th topic
$\mathbf{Q}_m^{(k)}$	the topic-specific latent vector ($\in \mathbb{R}^D$) for the m-th item *w.r.t.* the k-th topic
z_i ($z_{u,m}$)	latent topic associated with observation $i = \langle u, m\rangle$
θ_u	topic distribution ($\in \mathbb{R}^K$) of the u-th user
ϕ_k	item distribution ($\in \mathbb{R}^M$) of the k-th topic
α	Dirichlet priors over topics for topic models
β	Dirichlet priors over items for topic models
Ψ_0	Gaussian-Wishart priors for probabilistic matrix factorization

3.1 The Proposed Model

Our main idea is to construct clusters over items using topic models, and predict the rating using the ensemble of multiple topic-specific probabilistic matrix factorizations. Hence, our