Computational Creativity Based Video Recommendation

Wei Lu
Department of Computing
Hong Kong Polytechnic University
Kowloon, Hong Kong
cswlu@comp.polyu.edu.hk

Fu-lai Chung
Department of Computing
Hong Kong Polytechnic University
Kowloon, Hong Kong
korris.chung@polyu.edu.hk

ABSTRACT

Computational creativity, as an emerging domain of application, emphasizes the use of big data to automatically design new knowledge. Based on the availability of complex multi-relational data, one aspect of computational creativity is to infer unexplored regions of feature space and novel learning paradigm, which is particularly useful for online recommendation. Tensor models offer effective approaches for complex multi-relational data learning and missing element completion. Targeting at constructing a recommender system that can compromise between accuracy and creativity for users, a deep Bayesian probabilistic tensor framework for tag and item recommending is adopted. Empirical results demonstrate the superiority of the proposed method and indicate that it can better capture latent patterns of interaction relationships and generate interesting recommendations based on creative tag combinations.

Keywords

recommendation; serendipity; Bayesian tensor factorization

1. INTRODUCTION

Computational creativity, a study on algorithms for computer to generate artifacts that humans perceive to be creative, is a newly emerging field in today's e-commerce operations. Colton et al. [4] provide a definition for computational creativity research as: *"The philosophy, science and engineering of computational systems exhibit impersonally creative behaviors by taking on particular responsibilities."* Based on this definition, attempts including automatic culinary recipe generating system [9] have been made through cognitive flavor assessment. For recommendation, diversity also has become an important aspect for evaluation [11], where content-based systems usually suffer from over-specialization, since only items similar to those rated by users will be more likely recommended. Computational creativity is also related to serendipity, as defined in [6], a user-oriented measurement balancing between surprise and accuracy. Serendipitous recommendations by definition are also novel. Consider a recommender system that simply recommends movies directed by the user's favorite director, comparing to recommend a movie of the same director that the user was not aware of, a movie by a new director catering to the user's taste, is more likely to be not only novel but also serendipitous.

This whole idea of calculating computational creativity for recommender system from the automated processing of mass collection of item tags requires us to address several problems. First, we face the issue of effective user profiling. A two-dimensional collaborative filtering approach cannot be directly employed to build a tag-based recommender system, since it cannot efficiently capture the multi-dimensional characteristic and hence, will result in poorer recommendation performance. Second, creativity of new ideas built upon domain knowledge are difficult to be measured quantitatively. Besides traditional performance measurements, e.g. precision, recall, F-1, new evaluation metrics are required. Past approaches use historical viewing frequency based distance measurement, but as user preference is a complicated topic combination, pure frequencies would hardly express their real interest.

Fortunately, comparing to traditional user-item based recommender system, meta-data such as tags, which are reusable and shareable, play significant roles in helping manage online resources. As an additional source for item recognition, tags help in revealing user and improving user profiles that can be used in recommendation [5]. Hence, in this paper, we explore the following challenges: *given the inordinate amount of candidate items and their corresponding textual tag information now accessible online, is it possible to automatically generate serendipitous tag combinations for users via machine learning?* For example, given an online video recommender system, if a user has watched two videos with tag "Romance, Comedy" and "Thrilling, Horror", how would he react to a video tagged by "Comedy, Horror"? Will he find it creative or just as expected?

We assume that the historical behavior of users is a sound source for preference estimation, the high-level semantic topics of video tags can be regarded as a comprehensive representation of user features. There are some work on incorporating user features (e.g. age, gender, geographic information) and their social relationships (e.g. community groups) into a multi-way data analysis [10]. However, for video recommendation, users are more accurately clustered based on their overlapping of preferences, which can be represented by semantic relations among video tags. Integrating such multi-facet information, user profiles can be naturally modeled with higher-order data mining models. Tensor modeling is a well-known approach for representing latent relationships inherent in the multi-dimensional data.

There are typically three steps for tensor-based recommendation: (1) tensorial construction for multi-relational data; (2) tensor decomposition for latent feature (topic) representation; and (3) tensor reconstruction for interaction regeneration. Traditional multi-way factor models suffer from the drawback of failing to capture

SIGIR '16, July 17-21, 2016, Pisa, Italy
© 2016 ACM. ISBN 978-1-4503-4069-4/16/07...$15.00
DOI: http://dx.doi.org/10.1145/2911451.2914707

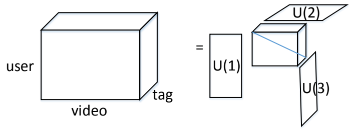


Figure 1: CP decomposition: a three mode user-video-tag relational dataset example. The tensor can be expressed in a matrix format: $\mathcal{X} = \{\lambda, U^{(1)}, U^{(2)}, \ldots, U^{(k)}\}$

coupled and nonlinear interactions between entities [13]. Also, they are not robust to datasets containing noisy and missing values. Through proper generative model, nonparametric Bayesian multi-way analysis algorithms are especially appealing, since they provide efficient ways to deal with distinct data types as well as data with missing values and noises. Moreover, since we are expected to recommend items not only considering the accuracy but also the creativity and serendipity, the posterior likelihood of Bayesian model can also be leveraged as a probabilistic ranking generating mechanism. Meanwhile, deep networks prove great empirical success in various domains [2]. They are capable of providing more compact nonlinear representations for feature learning. It would be interesting to adopt deep learning in one or more of the tensor modes and assess its effectiveness on tensor completion.

In this paper we address the aforementioned challenges through a novel Deep Tensor for Probabilistic Recommendation (DTPR) method. It breaks the task at hand into the following components: 1. a tensor construction stage of building user-item-tag correlation; 2. a tensor decomposition stage learning factors for each component mode; 3. a stage of tensor completion, which computes the creativity value of tag pairs; and 4. a recommender stage that ranks the candidate items according to both precision and creative consideration. This approach is evaluated using a real world video recommender system, with large amount of users, videos and corresponding video description tags.

2. PROPOSED METHODS

2.1 Frame of the Method

For a typical online video recommender system, the first step is to construct a third-order tensor from the video clicking logs; secondly, factoring the tensor using a decomposition method; and then reconstructing the decomposed elements. Following these steps, a three-way tensor, Tag × User × Video, is then constructed populated with the times of video being clicked by a user (Figure 1). For the factorization part, a Deep Canonical PARAFAC Factorization (DCPF) Model [8] is applied.

Accordingly, the next step is ranking the candidate items for recommendation. To encourage creative recommendation, it includes the consideration of tag pair preferences from the reconstructed tensor through introducing Bayesian surprise, which will be further elaborated in the following subsections. The preferences of users to rank the top-N list of item candidates are then probabilistically calculated.

2.2 Model Description

We briefly review the DCPF model discussed in [8]. Let $\mathcal{Y}$ denote an observed incomplete K-order tensor. For different types of observation, a noise function f can be applied on, depending on the data type being modeled [14] (e.g. Gaussian for real-valued data). The parameters of CP decomposition, $\{U^{(k)}\}_{k=1}^{K}$, are further represented by lower rank component factor score $V^{(k)}$ and factor loading $D^{(k)}$. $B^{(k)}$ is an indicator of which factors are actually used. Since it is fully conjugate, the sampler cycling steps are presented in Algorithm 1. In the algorithm, R is a positive integer indicating the rank of the core tensor. λ_r is the weight associated with the r^{th} rank component. a and b are hyper-parameters for δ_r. All the parameters with "ˆ" in the updating steps are the posterior expectation and precision. Details about the distributions are elaborated in [8].

Algorithm 1 Gibbs sampler steps

Begin
Initialize $U^{(k)}$, $B^{(k)}$, $V^{(k)}$, $D^{(k)}$
for iteration i **do**
  for each diagonal element r of the core tensor Λ, which is independently drawn from a normal distribution **do**
    1. Sample its independent conditionally conjugate posteriors from $\lambda_r \sim \mathcal{N}(\hat{\mu_r}, \hat{\tau_r}^{-1})$
    2. Update $\delta_r \sim Ga(a + \frac{1}{2}(R - r + 1), b + \frac{1}{2}\sum_r \tau_r^2 \delta_r)$
  end for
  for each mode factor matrix $U^{(k)}$ **do**
    1. Update binary indicator matrix $B^{(k)}$, factor loading matrix $V^{(k)}$ and factor score matrix $D^{(k)}$
    2. Sample column vector conditionally conjugate posteriors from $U_{i_k}^{(k)} \sim \mathcal{N}(\hat{\mu}_{i_k}^{(k)}, \hat{\Sigma}_{i_k}^{(k)-1})$
    3. Update hyper-parameters
  end for
end for

2.3 Calculation of Computational Creativity

In dynamic situations, surprise is an important indicator of the belief changing significance [1]. It can be also calculated as a quantization for creativity measure. According to Bayesian theorem, for an observer with prior distribution $P(X)$, the collection of data D leads to a re-evaluation of beliefs $P(X|D)$, the posterior distribution. Surprise is to define the distance between the prior and posterior distributions, specifically as:

$$S(D, X) = d[P(X), P(X|D)] \tag{1}$$

where d is a distance measure function. For example, if we use the relative entropy or Kullback-Liebler [7] divergence, which is a common practice for many well-know dissimilarity measures, the surprise S can be calculated as:

$$S(D, X) = logP(D) - \int_X P(X)logP(D|X)\, \mathrm{d}X \tag{2}$$

Since we can regard the novel tag generation process as a tensor completion problem, based on the posterior distribution of factor matrix $U^{(1)}$, the posterior distribution for co-occurring tags t_1, t_2 for user u can be written as:

$$p(y_{i_1=\{t_1,t_2\}, i_2=u, i_3}) = \prod_{i_3=1}^{n_3} \prod_{i_1} \mathcal{N}(y_{\mathbf{i}}|\hat{x}_{\mathbf{i}}, \hat{\tau}_e^{-1}) \tag{3}$$

where $\hat{x}_{\mathbf{i}} = \hat{p}_{i_k r}^{(k)} U_{i_k r}^{(k)} + \hat{q}_{i_k r}^{(k)}$, $\hat{p}_{i_k r}$ and $\hat{q}_{i_k r}^{(k)}$ are the posterior values calculated from the conditional posterior of $U_{i_k r}^{(k)}$.

2.4 Probabilistic Ranking Incorporated with Bayesian Surprise

Most existing tensor-based recommendation approaches rank the recommendations based on the values of reconstructed tensor. For example, Zheng et al. [15] utilize the linear add-up of the i^{th} row of each mode factor matrix. The higher weight it is, the more relevant user i is related to the topic. Since the goal of our task is not only providing accurate but also creative recommendations, instead of solely using the final reconstructed tensor from the model, we propose to rank the result of tensor modeling incorporating the Bayesian surprise value for generating the top-N list of candidate items to the users.

Using the reconstructed tensor $\hat{\mathcal{Y}}$, for each user u, two candidate lists can be created: (1) a list of items that the user might be interested in based on the posterior add-up values of user and item factor matrices; and (2) a list of tag preference pairs based on the normalized maximum values of Bayesian surprise metric. Assume the number of candidate item is n, and the number of tag pairs is J. J can be adjusted depending on how much compromise on creativity the system wishes to make. The score for ranking user-item pair is calculated as follows:

$$\mathcal{S}core_{u,v} = w_{u,v} \sum_{i=1}^{N_j \in v} \frac{S_{u,i}}{\sum_j^J S_{u,j}} \tag{4}$$

Where $\mathbf{w} = \sum_{r=1}^{R} \lambda_r \mathbf{u}_r^{(1)} \mathbf{u}_r^{(2)T}$ is the weight matrix obtained from reconstructed tensor, N_j is the number of tag pairs belonging to video v, in the top-J surprising list to user u. $\mathbf{S}$ is the surprise degree matrix of users to tag pairs. Each row vector of the final score matrix is then ranked in descending order, indicating the top-N weighted items as recommendation candidates.

3. EXPERIMENTAL RESULTS

3.1 Datasets

We perform experiments on a large-scale real-word dataset to verify the performance of the DTPR model on expressing high-level semantic features of user behavior and incorporating creativity consideration into recommendation. The real-world records were collected in three consecutive weeks at the same time slot from Tencent QQ browser (http://video.browser.qq.com) in August, 2014. The database stores historical clicking logs of active users extracted from the Hadoop distributed file system, with each record referring to one video clicked. Besides the time slot, user id, video id, number of clicks by this particular user and the total number of clicks in history, there are four categories of information provided as video description tag (in Chinese): (1) type (e.g. action, mystery); (2) region (e.g. US, main land China); (3) director; and (4) actor. Based on these information, a three-way User× Video × Tag tensor is constructed. There are $4,071,811$ samples in total with $137,465$ unique users, and $9,393$ unique videos they clicked. The rank of the tensor equals to 140.

3.2 Tag Completion Evaluation

Recall that our goal is to automatically generate serendipitous tag combinations for users, it is important to, analyze the tag completion performance part of the proposed method first. If we regard each type of tags as semantic words from different domains, this problem can also be regarded as multi-task clustering. We would like to evaluate how well the system ranks the correct items for each user based on single category and multiple categories of tags. Two metrics are employed for this: (1) the Mean Reciprocal Rank (MRR) and (2) Precision at 1 (P@1). MRR computes the inverse rank of the correct item and averages the score across the whole data. P at 1 computes the percentage of times the ground truth item are ranked as the top one.

From Table 1, although having high frequency of occurrence, tags from category II "Region" have the lowest MRR and P@1 due to its low differentiable vocabulary. For single layer DTPR, comparing to use purely actor tags, combining information from all these four categories increases the MRR by 1.7% (from 0.3662 to 0.3726). The 2-layer DTPR, in comparison, increases the MRR from 0.4110 to 0.3744, which is a 9% relative increase. Comparing to single layer DTPR, the increase in P@1 is also more obvious (36%) comparing to 1-layer implementation. This encouraging results indicate that a deep decomposition utilizing information from multiple aspects for factor matrix can better capture the semantic representation of user behavior.

Table 1: Results for 1-layer and 2-layer DTPR on different categories of tags.

	1-layer DTPR		2-layer DTPR	
	MRR	P@1	MRR	P@1
I	0.1867	0.0093	0.1974	0.0099
II	0.0009	0.0013	0.0009	0.0012
III	0.2900	0.2500	0.2990	0.2541
IV	0.3662	0.2233	0.3744	0.2350
ALL	0.3726	0.2249	0.4110	0.3197

3.3 Bayesian Surprise Assessment

For a subset of 3-way tensor, since Bayesian surprise is user dependent, the surprise value for each pair-wise combination of tags is presented in Figure 3. To make the figure clear, only the first $4,000$ pairs from a total $22,155$ are shown. A higher value indicates a more diverse posterior belief. In other words, if the semantic themes from the pair of tags co-occur in the same video, a higher Bayesian surprise would indicate that the user regard it as more creative. As indicated from the figure, the response of users to these $4,000$ tag pairs can generally be grouped into five folds. Within each fold, the surprise degree of each user varies according to the past video preference. The clustering of tag pairs also illustrates that the user preference has its probabilistic topic features. Hence comparing to traditional serendipity measurements, which are based on fixed distance of observed records, the posterior likelihoods are more suitable for creativity evaluation, as they incorporate the uncertainty and the cluster characteristics of user preference.

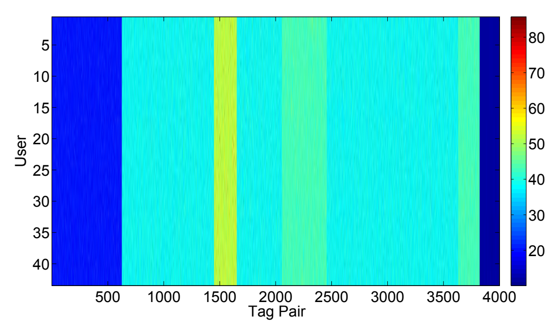
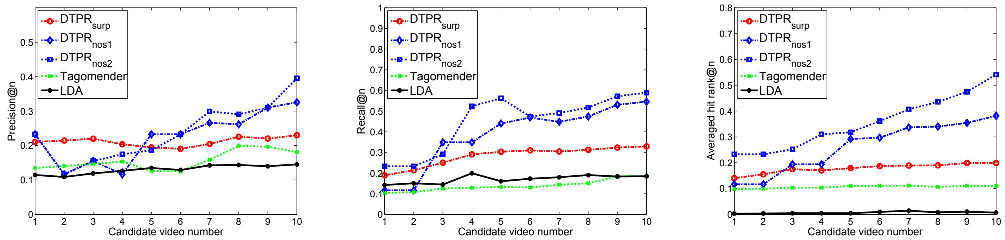


Figure 3: The surprise value for each user in the subset if the two tags occur together. X-axis indicates 4000 distinct tag combinations. Y-axis is 43 users

3.4 Performance Comparison

In this sub-section, we compare the recommendation performances with and without incorporating Bayesian surprise, using common evaluation metrics, including precision, recall and averaged hit rate.



(a) Precision@N (b) Recall@N (c) Averaged hit rate

Figure 2: Video recommendation performance comparison for single layer DTPR incorporated with Bayesian surprise (DTPR$_{surp}$), 1-layer (DTPR$_{nos1}$) and 2-layer (DTPR$_{nos2}$) DTPR without Bayesian surprise, Latent Dirichlet Allocation [3], and Tag matching method [12], with varying candidates number n ranging from 1-10.

The measures are defined as: $Precision = \frac{N_r}{N}$, $Recall = \frac{N_r}{N_t}$, $Averaged\ hit\ rate = \frac{\sum_{i=1}^{N_r} \frac{1}{r_i}}{N_t}$ where N_r is the number of relevant videos in top-N item candidates, N_t is the ground truth of relevant videos that user has watched, r_i is the rank of each relevant video in the recommender pool. The pair number J is fixed at 5.

As shown in Figure 2, we compare the proposed model in single layer and 2-layer version with and without incorporating Bayesian surprise for item ranking. The increase of pool number will enhance the probability to recommend the favored video. Thus, the three metrics have the trend of gradually growing. Although single layer and 2-layer DTPR have similar precision for predicting, a multi-layer implementation allows an obvious higher hit rate, which indicates that it can pick the correct choice of users at an earlier stage of recommendation. Through incorporating Bayesian surprise, the performance does degrade due to the additional uncertainty it introduces to the item ranking. However, it still outperforms both LDA and Tagommender [12] method. Considering the surprise it brings to the users, which will in return enhance user experience, this kind of recommending has its merits in practice.

4. CONCLUSION

Given the increasing growth in large-scale multi-relational data, we study the computational creativity problem in video recommendation domain. To effectively learn user-item-tag correlations, we utilize deep Bayesian tensor model, which provides an effective way for joint analysis of user and video features. Through a scalable framework for tensor decomposition and completion, and through introducing Bayesian surprise into probabilistic ranking, we are able to recommend personalized items taking creativity as a consideration. Our model can perform fully conjugate Bayesian inference via Gibbs sampling inference. The quantization index, Bayesian surprise, for computational creativity is assessed. Currently the creativity is valuated based on generating new combination of existent items. Creative construction for data from previously unexplored domain based on current knowledge are also appealing for future targets.

5. ACKNOWLEDGMENTS

This research was supported by project B-Q43W, B-Q49B, and G-UA3W. We would also like to thank Kunfeng Lai and the Mobile Internet group at Tencent Shenzhen for providing the data.

6. REFERENCES

[1] P. Baldi and L. Itti. Of bits and wows: a bayesian theory of surprise with applications to attention. *Neural Networks*, 23(5):649–666, 2010.

[2] Y. Bengio. Learning deep architectures for ai. *Foundations and trends® in Machine Learning*, 2(1):1–127, 2009.

[3] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. *the Journal of machine Learning research*, 3:993–1022, 2003.

[4] S. Colton, G. A. Wiggins, et al. Computational creativity: the final frontier? 2012.

[5] N. Ifada. A tag-based personalized item recommendation system using tensor modeling and topic model approaches. In *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval*, pages 1280–1280. ACM, 2014.

[6] P. B. Kantor, L. Rokach, F. Ricci, and B. Shapira. *Recommender systems handbook*. Springer, 2011.

[7] S. Kullback. *Information theory and statistics*. Courier Corporation, 1997.

[8] W. Lu and F.-l. Chung. Deep bayesian tensor for recommender system. In *Proceedings of the ECMLPKDD 2015 Doctoral Consortium*, pages 135–144, 2015.

[9] F. Pinel and L. R. Varshney. Computational creativity for culinary recipes. In *CHI'14 Extended Abstracts on Human Factors in Computing Systems*, pages 439–442. ACM, 2014.

[10] P. Rai, Y. Wang, and L. Carin. Leveraging features and networks for probabilistic tensor decomposition. 2015.

[11] F. Ricci, L. Rokach, and B. Shapira. *Introduction to recommender systems handbook*. Springer, 2011.

[12] S. Sen, J. Vig, and J. Riedl. Tagommenders: connecting users to items through tags. In *Proceedings of the 18th international conference on World wide web*, pages 671–680. ACM, 2009.

[13] Z. Xu, F. Yan, et al. Infinite tucker decomposition: Nonparametric bayesian models for multiway data analysis. *arXiv preprint arXiv:1108.6296*, 2011.

[14] Z. Xu, F. Yan, and Y. Qi. Bayesian nonparametric models for multiway data analysis. 2013.

[15] N. Zheng, Q. Li, S. Liao, and L. Zhang. Flickr group recommendation based on tensor decomposition. In *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval*, pages 737–738. ACM, 2010.