# Multi-Modal Multi-Correlation Person-Centric News Retrieval

Zechao Li[1], Jing Liu[1], Xiaobin Zhu[1], Hanqing Lu[1,2]
[1]Institute of Automation, Chinese Academy of Sciences, Beijing, China, 100086
[2]China-Singapore Institute of Digital Media, 21 Heng Mui Keng Terrace, Singapore, 119613
{zcli, jliu, xbzhu, luhq}@nlpr.ia.ac.cn

## ABSTRACT

In this paper, we propose a framework of multi-modal multi-correlation person-centric news retrieval, which integrates news event correlations, news entity correlations, and event-entity correlations simultaneously by exploring both text and image information. The proposed framework is confined to a person-name query and enables a more vivid and informative person-centric news retrieval by providing two views of result presentation, namely a query-oriented multi-correlation map and a ranking list of news items with necessary descriptions including news image, news title and summary, central entities and relevant news events. First, we pre-process news articles using natural language techniques, and initialize the three correlations by statistical analysis about events and entities in news articles and face images. Second, a Multi-correlation Probabilistic Matrix Factorization (MPMF) algorithm is proposed to complete and refine the three correlations. Different from traditional Probabilistic Matrix Factorization (PMF), the proposed MPFM additionally considers the event correlations and the entity correlations as well as the event-entity correlations during the factor analysis. Third, the result ranking and visualization are conducted to present search results relevant to a target news topic. Experimental results on a news dataset collected from multiple news websites demonstrate the attractive performance of the proposed solution for news retrieval.

## Categories and Subject Descriptors

H.3.3 [**Information Systems**]: Information Storage and Retrieval - *Information Search and Retrieval*

## General Terms

Algorithms, Experimentation

## Keywords

Factor analysis, Multi-modal, Multi-correlation, Person-centric, Probabilistic matrix factorization


*CIKM'10,* October 26–30, 2010, Toronto, Ontario, Canada.



**Figure 1: An illustrate example of connections among news entities and news events.**

## 1. INTRODUCTION

Along with information technology development and internet globalization, online news articles have enjoyed explosive growth and received high popularity from over half of web users. Such spurting growth urges the necessity for efficient organizing the large amount of news articles from multiple new sources.

A news article is defined as a specific event arose by specific people or an organization which happens at a certain time and place. That is, a news article corresponding to a specific news event can be identified according to the following '4W' elements: Who (person or organization), When (time), Where (locations), and What (event). In particular, 'Who' as a news entity should be paid special attention because the social network among different persons can be viewed as a kind of indirect connection among news events as well as their textual relevance. Accordingly, news event and news person should be considered as the two basic items in news retrieval, special for person-centric news retrieval in this paper. The both items are correlated to each other. Specifically, different news articles may be relevant when the news events happen on the same or related persons, and the persons appearing in the same news event should also be related by certain social interaction. As shown in Figure 1, the five news articles are related to some extent due to a shared person (*Yao Ming*), and different persons appearing in these articles are connected with specific social relation-

ships, such as *Yao Ming* is the teammate of *McGrady* and *Battier*, the husband of *Ye Li*, and the rival of *Howard*. It is an important and challenging problem to effectively explore the both items and their within- and inter- correlations to organize and search possible relevant news events on Internet, so as to enable an informative overview about a target news topic.

Some researchers have fixed their attention on exploring correlations within news events or news entities in various news-related applications. However, most of them depend on the text analysis and neglect the inter-correlation mining between event and entity. Usually, news web pages contain news images to vividly describe a specific news event, in which central news actors (persons or an organization) and locations maybe appear. Thus, the importation of news images as well as textual details is valuable to deeply understand news articles, and to describe them more precisely. Currently, few work attempt to employ the multi-modal analysis in news description, and jointly explore available correlations among events and entities to discover or correct some implicit ones during news retrieval.

In this paper, we propose a framework of Multi-modal Multi-correlation Person-centric News Retrieval (MMPNR) by integrating news event correlations, news entity correlations, and event-entity correlations simultaneously and seamlessly and employing both text information and image information. First, we pre-process news articles using common natural language techniques and face recognition technologies to obtain event correlations, entity correlations, and partial event-entity correlations respectively. Second, a complete event-entity correlation is estimated by a Multi-correlation Probabilistic Matrix Factorization (MPMF) model, which is an extended version of traditional Probabilistic Matrix Factorization (PMF) [23] with additional consideration of the with-in item correlations, i.e., the event correlation and the entity correlation. Third, the result ranking and visualization are conducted to present news searching results. Different from a typical news browser, which only presents a ranking list of relevant news items, we will additionally provide user a query-oriented multi-correlation map as an intuitive and concise resulting presentation. Experimental results on a news dataset collected from multiple sources demonstrate the effectiveness of the proposed MMPNR framework.

The rest of this paper is organized as follows. Section 2 presents a brief review of related work. Section 3 gives an overview of our news retrieval system. Section 4 presents the correlation initialization using multi-modal analysis. We introduce the proposed MPMF model to complete and refine the three correlations in Section 5. Section 6 gives the result ranking and visualization in MMPNR. Section 7 shows our evaluation of the methods using a large news article set and discusses our analysis of the results in detail. Section 8 concludes this paper.

## 2. RELATED WORK

The task of relation extraction has been traditionally studied as to extract predefined semantic relations between pairs of entities in text. The supervised methods [30, 11, 32, 10] require a set of labeled training examples of the predefined relations to learn an extractor. However, the labeled examples are scarce and expensive. The bootstrapping approaches [4, 1, 33, 8] relax the requirement on training data through iteratively discovering extraction patterns and identifying entity relations with a small number of seeds. Based on bootstrapping methods, some bootstrapping systems [4, 1, 33] have been explored. Brin [4] extracted many author-title relationships from the Web by matching phrases, supposing that a syntactic pattern corresponded to a relation. Similar to Brin's, the well-known Snowball [1] employed the pattern-entity duality to iteratively extract specific relationships from the Web. Taking Snowball as the basis, Zhu et al. [33] produced a system named StatSnowball, which significantly improved the retrieval performance on recall and precision. It used the general discriminative Markov logic networks, which subsumed logistic regression and conditional random fields, weights each pattern by maximum likelihood estimation, and can be configured to perform open information extraction (OpenIE) [2]. StatSnowball can perform joint inference by using the Markov logic networks, while the O-CRFs [3] treat sentences independently. Probabilistic relation model [9, 18] has been proposed to estimate the relations among entities. Relational techniques such as PRMs [9] extended generative methods to deal with various combinations of probabilistic dependency among entities. Such mehods can be computationally expensive, and may not scale to the large amount of data typically collected by social media websites. Some work focus on relational learning methods [18] through pairwise relationships among entities, which involve loss of information when data has high-order interactions.

Sekine et al. [25] proposed 150 types of named entity, which were useful in information extraction and Question and Answering (Q&A) in the newspaper domain. KnowItAll [7, 8, 26] is a system for automating the tedious process to extract large collections of facts from the web, which is based on generic extraction rules to generate candidates, co-occurrence statistic, and a naive Bayes classifier. It learned an effective pattern to extract relevant entities from relevant and irrelevant terms for expected entities. It found entity names in the same class as a given example by using several syntactic patterns. However, it required large numbers of search queries and downloaded web pages.

Event detection from news articles has been extensively studied. There are some approaches based on clustering [22, 29, 24]. Naughton et al. [22] discovered clusters at sentence level using hierarchical algorithms and regarded that each cluster pointed to an event with assumption that a news article could point to different events. Yang et al. [29] proposed an agglomerative clustering algorithm, in which the similarities between the incoming document and the known events were computed and a threshold was applied to make a decision. A new model was proposed in [24] to detect news events using a label based clustering approach. A unique thinking was proposed in [31], where the authors distinguished the concepts of relevance and redundancy. A probabilistic model [15] attempted to identify events within a corpus of historical news articles with the help of time information, user feedback, and content information.

Factor analysis [27] has been widely utilized in many fields [34, 12, 20, 21, 23, 19, 5, 16]. Zhu et al. [34] proposed a joint matrix factorization combining both linkage and document-term matrices to improve the hypertext classification. The content information and link strictures were seamlessly combined through a single set of latent factors. The discovered latent factors (bases) explained both content information and link structures, and were used to classify the web

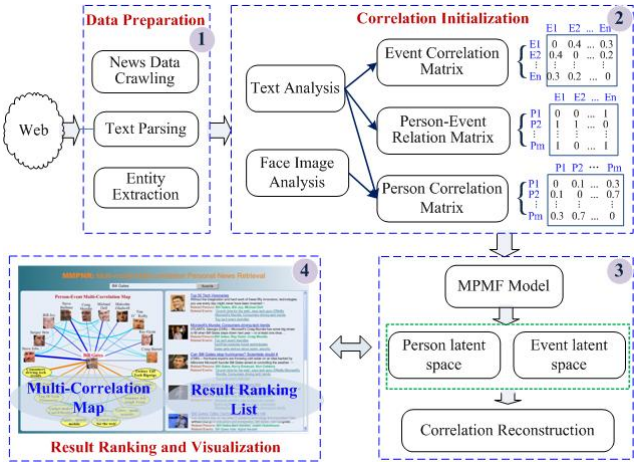


**Figure 2: The framework of the proposed MMPNR.**

pages. A variety of probabilistic factor-based models have been proposed [12, 20, 21], which can be viewed as graphical models in which latent factor variables have directed connections to variables that represent user ratings. The major drawback is that potentially slow or inaccurate approximations are required for the estimation of the posterior distribution over latent factors. Probabilistic Matrix Factorization (PMF) [23] was proposed to find only point estimates of model parameters and hypermeters, instead of inferring the full posterior distribution over them. It models the user-item matrix as a product of two low-rank user and item matrices. The computation cost of PDF is linear with the number of observations. A social recommendation has been proposed based on the probabilistic matrix factorization model, named as SoRec [19]. It fused the user-item matrix with the users' social trust networks by sharing a common latent low-dimensional user feature matrix. A probabilistic polyadic factor model [5] was proposed to analyze multiple-dimensional data such as networked data in social networks and directly model all the dimensions simultaneously in a unified framework. In order to discover community structure from various social contexts and interactions, MetaFac (MetaGraph Factorization) [16] proposed an efficient factorization approach through analyzing time-varying and multi-relational data.

To make full use of the within- and inter- correlations among events and entities, in this paper, we propose a MPMF model to learn the person latent space and the event latent space by exploring the three correlation matrices simultaneously and seamlessly. Based on the reconstructed correlations over the latent spaces, we provide users a concise and informative result browsing with a query-oriented correlation map and a ranking list of news items. As far as we know, the proposed MPMF-based news retrieval is the first one attempting to mine the within- and inter- correlations among events and entities in a simultaneous and seamless form.

# 3. OVERVIEW OF OUR SYSTEM

In this section, we will briefly overview the framework of the proposed MMPNR (shown in Figure 2), which includes four components, namely data preprocessing, correlation initialization, correlation reconstruction, and result ranking and visualization.

First, we collect and preprocess news data. A large scale of news articles are crawled from some distinguished news sites including ABCNews.com[1], BBC.co.uk[2] and CNN.com[3]. We first parse these news articles into news titles, summaries, texts, URLs and images of news pages. Necessary text preprocessing including word separation and stop-words filtering are conducted. Then we extract news entities (Time, Person or Organization, and Location) according to [25]. In this paper, we view Person (or Organization)as entities, while the time and location are used to identify news events.

The second component introduced in Section 4 aims to initialize three kinds of correlations: the event correlation, the person correlation, and the person-event correlation. In particular, the event correlation is estimated via the TF-IDF model on text information (news title, summary, and details) from news web pages. For the person correlation, a linear combination of the two co-occurrences of person name entities within text information and faces on person images respectively. We utilize the occurrence of a specific person in an event to obtain a binary relationship between the person and the event.

The third component is the basic component in MMPNR, which is demonstrated in Section 5. We apply the multi-correlation probabilistic matrix factorization model to mine the hidden relations. We connect these three different correlations simultaneously and seamlessly through the shared person latent feature space and event latent feature space, that is, the person latent feature space in the news person relational matrix is the same in the person-event correlation matrix and the event latent feature space in the news event similarity matrix is the same in the person-event correlation matrix. By performing factor analysis via MPMF, the low-dimension person latent features and event latent features are learned, which can be used to reconstruct the news person-event correlations.

The fourth component described in Section 6 is the result ranking and visualization in MMPNR, which obtains and displays query-related search results to the end users. To give users vivid and informative organization of news results, we divide the user interface into two parts. The left part gives users a query-oriented relation graph, in which the relations between the query and events (or persons) are illustrated. In the right part, we present a ranking list of related news events with their titles, and the most relevant persons and events respectively.

# 4. CORRELATION INITIALIZATION

In this section, we will explain how to estimate the three correlations from multi-modal information on news web pages. The estimated correlations is viewed as the initialized inputs of the MPMF model, which will be introduced in Section 5. The details about the estimation are presented as follows.

## 4.1 Person-Event Correlation Matrix

As mentioned above, we employ the binary relationship to measure the person-event correlation $R$, that is, if a news person $i$ appears in a news event $j$, $R_{ij} = 1$ and $R_{ij} = 0$

[1]http://abcnews.go.com/
[2]http://www.bbc.co.uk/
[3]http://edition.cnn.com/

otherwise. Because the amount of online news articles is too large, the person-event relation matrix $R$ is very sparse, which is one of the reasons we employ the probabilistic matrix factorization model.

## 4.2 Event Correlation Matrix

From the aspect of utilizing the contents, TF-IDF is still the dominant technique to represent document, and cosine similarity is the generally used similarity metric. Therefore, we adopt the TF-IDF model and cosine similarity to measure the news event similarity matrix $S$. Considering the difference of the importance of news article's title, summary and text to a news event, we process them separately and linearly combine them. Besides, the information of title is the most important to the news event and the information of summary is more important than the information of text to the news event. In our experiments, we combine these three kinds of similarities as

$$S = \alpha \times S^{title} + \beta \times S^{summary} + (1 - \alpha - \beta) \times S^{text}, \quad (1)$$

where $S$, $S^{title}$, $S^{summary}$ and $S^{text}$ represent the similarity of event, title, summary and text, respectively.

## 4.3 Person Correlation Matrix

In view of current news web pages containing images and persons always appearing in images of news articles, we are supposed to utilize not only the text information, but also the information of images to calculate the co-occurrence of people in news events. Thus, we combine the text information and the image information to calculate the relationship among news persons. First, we use the formula $C_{iq}^{Text} = 2f(i,q)/(f(i)+f(q))$ to calculate the co-occurrence based on the textual information, where $f(i,q)$, $f(i)$ and $f(q)$ denote the number of news articles including person $i$ and person $q$ simultaneously, the count of news articles containing person $i$, and the count of news articles containing person $q$, respectively. We apply the face detection and matching methods to process the information of image, which will be explained in detail as follows. We employ the same formula to calculate the co-occurrence based on the image information $C_{iq}^{Img}$.

We first submit names of persons to Wikipedia[4], crawl and parse the corresponding returned web pages, and download images in the resume tables. And then we adopt the face detection approaches to detect the face parts in images.To determine whether a specific person appears in the news images or not, we adopt SIFT flow [17] to match the face part of the specific person's image from Wikipedia with any face part detected in the image from news web page. An illustrate example is given in Figure 3. The face parts (b), which are the face parts of images derived by submitting persons' names in (a), are used to be matched with the face parts in image (c) from news web pages. The matched results by SIFT flow algorithm are presented in (c). According to the matching results, we can derive the image co-occurrence matrix as shown in (d).

The SIFT flow approach assumes SIFT descriptors extracted at each pixel location are constant with respect to the pixel displacement field and allows a pixel in one image to match any other pixel in the other image. We still want to encourage smoothness of the pixel displacement field by

[4]http://en.wikipedia.org/wiki/Main_Page

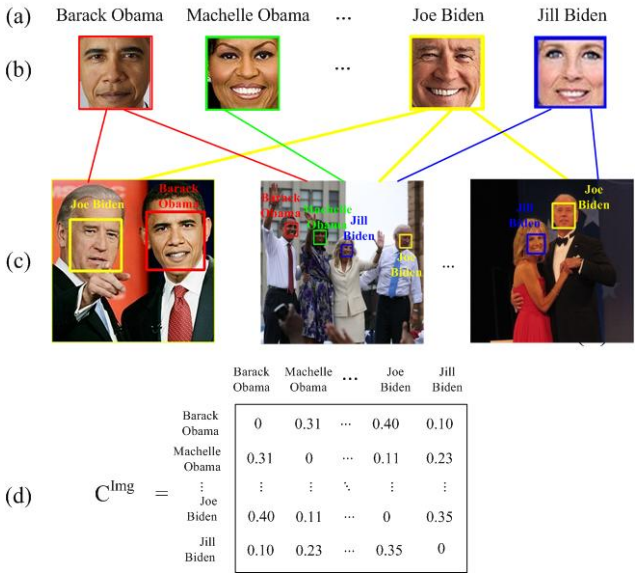


**Figure 3: An illustrate example of face detection and matching. (a) the names of news persons; (b) the faces detected in images crawled from Wikipedia by submitting names in (a); (c) the face matching results; (d) the co-occurrence matrix based on (c). The red, green, blue and yellow boxes represent "Barack Obama ", "Michelle Obama ", "Joy Biden" and "Jill Biden", respectively.**

encouraging close-by pixels to have similar displacements. It formulates the correspondence search as a discrete optimization problem on the image lattice with the following cost function

$$l(\mathbf{w}) = \sum_{\mathbf{p}} \|s_1(\mathbf{p}) - s_2(\mathbf{p}+\mathbf{w})\|_1 + \frac{1}{\sigma^2}\sum_{\mathbf{p}}(u_{\mathbf{p}}^2 + v_{\mathbf{p}}^2)$$
$$+ \alpha \sum_{(\mathbf{p},\mathbf{q})\in\varepsilon} \min(|u(\mathbf{p}) - u(\mathbf{q})|, \frac{d}{\alpha}) + \min(|u(\mathbf{p}) - u(\mathbf{q})|, \frac{d}{\alpha}),$$

where $\mathbf{w}(\mathbf{p}) = (u(\mathbf{p}), v(\mathbf{p}))$ is the displacement vector at pixel location $\mathbf{p} = (x, y)$, $s_i(\mathbf{p})$ is the SIFT descriptor extracted at location $\mathbf{p}$ in image $i$ and $q$ is the spatial neighborhood of a pixel. Parameters $\sigma = 300$, $\alpha = 0.5$ and $d = 2$ are fixed in our experiments. Based on the results of matching, we decide whether a person emerges in the news images. Finally, we obtain an indicator matrix with each column representing whether a news person appears in news images or not. We statistic term frequency and calculate the co-occurrence similar to text processing. We linearly integrate these two co-occurrences as follows:

$$C_{iq} = (1 - \gamma) \times C_{iq}^{Text} + \gamma \times C_{iq}^{Img}. \quad (2)$$

# 5. CORRELATION RECONSTRUCTION

Through the initializing component, we have accomplished the preparation for the following correlation reconstruction i.e., the complement and refinement of the initialized correlations. For clarity, we first introduce the standard PMF model. Then, we present our proposed MPMF model with additional consideration of the within correlations about events and entities respectively.

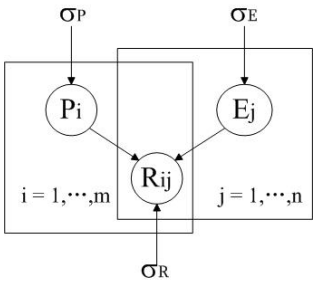


**Figure 4: Graphical model for Probabilistic Matrix Factorization (PMF).**

## 5.1 PMF Model

PMF model [23] was proposed to handle very large, sparse, and imbalanced dataset in collaborative filtering, based on the assumption that users who have rated similar sets of movies are likely to have similar preferences. Following, let us take collaborative filtering for example to illustrate the probabilistic matrix factorization model.

In order to learn the characteristic of the users, matrix factorization is employed to factorize the user-item matrix. Suppose we have $m$ users, $n$ movies, and the rating values within the interval [0, 1]. If the rating values are integer from 1 to $Q$ (a bigger integer value more than 1, such as 3 and 5), we can map the ratings $1, ..., Q$ to the internal [0, 1] using the function $h(x) = (x-1)/(Q-1)$. Let $R_{ij}$ denote the rating of user $i$ for movie $j$, $R \in \mathbb{R}^{m \times n}$ denote the rating matrix. The basic idea of probabilistic matrix factorization is to derive two high-quality $d$-dimensional ($d$ is lower than $\min(m, n)$.) latent feature spaces $P \in \mathbb{R}^{d \times m}$ and $E \in \mathbb{R}^{d \times n}$, which denote the latent user and movie feature spaces respectively. The column vectors $P_i$ and $E_j$ represent user-specific and movie-specific latent feature vectors, which are not unique. A probabilistic model with Gaussian observation noise is employed as shown in Figure 4, and the conditional distribution over the observed rating is defined as:

$$p(R|P, E, \sigma_R^2) = \prod_{i=1}^{m} \prod_{j=1}^{n} [\mathcal{N}(R_{ij}|g(P_i^T E_j), \sigma_R^2)]^{I_{ij}}, \quad (3)$$

where $\mathcal{N}(x|\mu, \sigma^2)$ denotes the probabilistic density function, in which the conditional distribution is defined as the Gaussian distribution with mean $\mu$ and variance $\sigma^2$, and $I_{ij}$ is the indicator function that is equal to 1 if user $i$ rated movie $j$ and equal to 0 otherwise. The function $g(x)$ is a logistic function defined as $g(x) = 1/(1 + exp(-x))$, which makes it possible to bound the range of $P_i^T E_j$ within the interval [0, 1]. As described in [6, 28], zero-mean spherical Gaussian priors are placed on user and movie feature vectors:

$$p(P|\sigma_P^2) = \prod_{i=1}^{m} \mathcal{N}(P_i|0, \sigma_P^2 \mathbf{I}), \quad (4)$$

$$p(E|\sigma_E^2) = \prod_{j=1}^{n} \mathcal{N}(E_j|0, \sigma_E^2 \mathbf{I}), \quad (5)$$

where $\mathbf{I}$ is an identity matrix.

Through a Bayesian inference, the posterior distribution over the user and movie features is given by:

$$\begin{aligned} p(P,E|R, \sigma_R^2, \sigma_P^2, \sigma_E^2) &\propto p(R|P, E, \sigma_R^2) p(P|\sigma_P^2) p(E|\sigma_E^2) \\ &= \prod_{i=1}^{m} \prod_{j=1}^{n} [\mathcal{N}(R_{ij}|g(P_i^T E_j), \sigma_R^2)]^{I_{ij}} \\ &\times \prod_{i=1}^{m} \mathcal{N}(P_i|0, \sigma_P^2 \mathbf{I}) \times \prod_{j=1}^{n} \mathcal{N}(E_j|0, \sigma_E^2 \mathbf{I}). \end{aligned} \quad (6)$$

Thus, we can derive the log of the posterior distribution given by Eq. 6, described as:

$$\begin{aligned} &\ln p(P, E|R, \sigma_R^2, \sigma_P^2, \sigma_E^2) = \\ &- \frac{1}{2\sigma_R^2} \sum_{i=1}^{m} \sum_{j=1}^{n} I_{ij}^R (R_{ij} - g(P_i^T E_j))^2 \\ &- \frac{1}{2\sigma_P^2} \sum_{i=1}^{m} P_i^T P_i - \frac{1}{2\sigma_E^2} \sum_{j=1}^{n} E_j^T E_j \\ &- \frac{1}{2} (\sum_{i=1}^{m} \sum_{j=1}^{n} I_{ij}^R \ln \sigma_R^2 + md \ln \sigma_P^2 + nd \ln \sigma_E^2) + \mathcal{C}, \end{aligned} \quad (7)$$

where $\mathcal{C}$ is a constant that does not depend on the parameters. Maximizing the log-posterior distribution over user and movie features given by Eq. 7 with hyperparameters (i. e. the observation noise variance and prior variances) kept fixed is equivalent to minimizing the following sum-of-squared-errors objective functions with quadratic regularization terms:

$$\begin{aligned} L(P, E) = &\frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{n} I_{ij}^R (R_{ij} - g(P_i^T E_j))^2 \\ &+ \frac{\lambda_P}{2} \|P\|_F^2 + \frac{\lambda_E}{2} \|E\|_F^2, \end{aligned} \quad (8)$$

in which $\lambda_p = \sigma_R^2/\sigma_P^2$, $\lambda_E = \sigma_R^2/\sigma_E^2$, and $\|\cdot\|_F^2$ denotes the Frobenius norm. Eq. 8 can be solved using gradient methods, such as the conjugate gradient, quasi-Newton methods and steepest descent method. Through performing gradient descent in $P$ and $E$ as described in Eq. 9 and Eq. 10, we can find a local minimum of the objective function given by Eq. 8.

$$\frac{\partial L}{\partial P_i} = \sum_{j=1}^{n} I_{ij}^R g'(P_i^T E_j)(g(P_i^T E_j) - R_{ij}) E_j + \lambda_P P_i \quad (9)$$

$$\frac{\partial L}{\partial E_j} = \sum_{i=1}^{m} I_{ij}^R g'(P_i^T E_j)(g(P_i^T E_j) - R_{ij}) P_i + \lambda_E E_j \quad (10)$$

$g'(x)$ is the derivative of logistic function $g'(x) = exp(x)/(1 + exp(x))^2$.

The experimental results in [23] demonstrate that probabilistic matrix factorization performs very well on the very large, sparse and imbalanced dataset and takes time linear in the number of observations using steepest descent. The graphical model shown in Figure 4 represents the method how to derive the users latent feature spaces based on the user-item matrix without considering the users' social network and the items' similarities. In the next subsection, we will propose an algorithm to integrate the user-item matrix, users' social network and items' similarity matrix simultaneously and seamlessly.

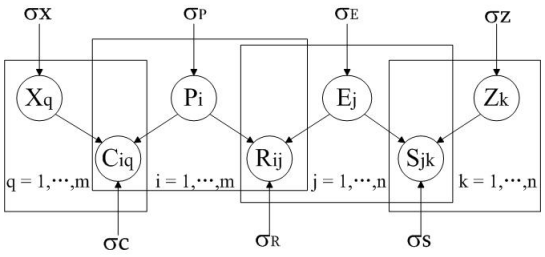


**Figure 5: Graphical model for MPMF.**

## 5.2 MPMF Model

In the problem of our person-centric news retrieval, we consider two issues: news person and event. We can have three kinds of relations: person-event correlation, person correlation and event correlation. The person-event relation in the problem of person-centric news retrieval can be analogous to user-item relation in recommender system. Furthermore, due to the fast explosion of online news articles, the correlations among news persons and news events are usually very sparse. The probabilistic matrix factorization algorithm as a natural and feasible option is employed to conduct our work. However, the standard probabilistic matrix factorization model can only employ one relation. Then we extend the model to integrate the news person correlation and news event correlation, named as Multi-correlation Probabilistic Matrix Factorization (MPMF). We employ the probabilistic factor analysis to factorize person-event correlation matrix, person correlation matrix and event correlation matrix, and connect these three different data resources through the shared person latent feature space, that is, the person latent feature space in the person-event correlation matrix is the same in the person correlation space, and the shard event latent feature space, that is, the event latent feature space in the person-event correlation matrix is the same in the event correlation matrix.

To learn the person and event latent feature spaces, we model our problem using the graphical model described in Figure 5. Suppose we have $m$ persons and $n$ events. Let $R \in \mathbb{R}^{m\times n}$, $C \in \mathbb{R}^{m\times m}$ and $S \in \mathbb{R}^{n\times n}$ denote the person-event relation matrix, person correlation matrix and event similarity matrix respectively. Let $R_{ij}$ represent the relation of person $i$ and event $j$ within the range [0, 1], $C_{iq} \in [0, 1]$ denote the relation between person $i$ and person $q$, and $S_{jk} \in [0, 1]$ denote the similarity between event $j$ and event $k$. Let $P \in \mathbb{R}^{d\times m}$, $E \in \mathbb{R}^{d\times n}$, $X \in \mathbb{R}^{d\times m}$ and $Z \in \mathbb{R}^{d\times n}$ be person, event, person factor and event factor latent feature matrices, with column vectors $P_i$, $E_j$, $X_q$ and $Z_k$ representing person-specific, event-specific, person factor-specific and event factor-specific latent feature vectors, respectively.

The probabilistic model with Gaussian observation noise is adopted and the conditional distributions are defined as:

$$p(R|P,E,\sigma_R^2) = \prod_{i=1}^{m}\prod_{j=1}^{n}[\mathcal{N}(R_{ij}|g(P_i^T E_j),\sigma_R^2)]^{I_{ij}^R}, \quad (11)$$

$$p(C|P,X,\sigma_C^2) = \prod_{i=1}^{m}\prod_{q=1}^{m}[\mathcal{N}(C_{iq}|g(P_i^T X_q),\sigma_C^2)]^{I_{iq}^C}, \quad (12)$$

$$p(S|E,Z,\sigma_S^2) = \prod_{j=1}^{n}\prod_{k=1}^{n}[\mathcal{N}(S_{jk}|g(E_j^T Z_k),\sigma_S^2)]^{I_{jk}^S}, \quad (13)$$

where $I_{ij}^R$ is the indicator function that is equal to 1 if the relation between news person $i$ and news event $j$ is more than 0 and equal to 0 otherwise. $I_{iq}^C$ and $I_{jk}^S$ are defined similarly.

We also place zero-mean spherical Gaussian priors on person, event, person factor and event factor feature vectors.

$$p(P|\sigma_P^2) = \prod_{i=1}^{m}\mathcal{N}(P_i|0,\sigma_P^2\mathbf{I}) \quad (14)$$

$$p(E|\sigma_E^2) = \prod_{j=1}^{n}\mathcal{N}(E_j|0,\sigma_E^2\mathbf{I}) \quad (15)$$

$$p(X|\sigma_X^2) = \prod_{q=1}^{m}\mathcal{N}(X_q|0,\sigma_X^2\mathbf{I}) \quad (16)$$

$$p(Z|\sigma_Z^2) = \prod_{k=1}^{n}\mathcal{N}(Z_k|0,\sigma_Z^2\mathbf{I}) \quad (17)$$

Hence, similar to Eq. 7, through a simple Bayesian inference, we can obtain the log of the posterior distribution:

$$\begin{aligned}
\ln p(P,E,X,Z|R,C,S,\sigma_R^2,\sigma_C^2,\sigma_S^2,\sigma_P^2,\sigma_E^2,\sigma_X^2,\sigma_Z^2) = \\
&-\frac{1}{2\sigma_R^2}\sum_{i=1}^{m}\sum_{j=1}^{n}I_{ij}^R(R_{ij}-g(P_i^T E_j))^2 \\
&-\frac{1}{2\sigma_C^2}\sum_{i=1}^{m}\sum_{q=1}^{m}I_{iq}^C(C_{iq}-g(P_i^T X_q))^2 \\
&-\frac{1}{2\sigma_S^2}\sum_{j=1}^{n}\sum_{k=1}^{n}I_{jk}^S(S_{jk}-g(E_j^T Z_k))^2 \\
&-\frac{1}{2\sigma_P^2}\sum_{i=1}^{m}P_i^T P_i-\frac{1}{2\sigma_E^2}\sum_{j=1}^{n}E_j^T E_j \\
&-\frac{1}{2\sigma_X^2}\sum_{q=1}^{m}X_q^T X_q-\frac{1}{2\sigma_Z^2}\sum_{k=1}^{n}Z_k^T Z_k \\
&-\frac{1}{2}((\sum_{i=1}^{m}\sum_{j=1}^{n}I_{ij}^R)\ln\sigma_R^2+(\sum_{i=1}^{m}\sum_{q=1}^{m}I_{iq}^C)\ln\sigma_C^2) \\
&-\frac{1}{2}\sum_{j=1}^{n}\sum_{k=1}^{n}\mathbf{I}_{jk}^S\ln\sigma_S^2-\frac{1}{2}md\ln\sigma_P^2 \\
&-\frac{1}{2}(nd\ln\sigma_E^2+md\ln\sigma_X^2+nd\ln\sigma_Z^2)+\mathcal{C}. \quad (18)
\end{aligned}$$

As described above, the equivalent optimization problem is to minimize the following objective function:

$$\begin{aligned}
L(P,E,X,Z) = &\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{n}I_{ij}^R(R_{ij}-g(P_i^T E_j))^2 \\
&+\frac{\lambda_C}{2}\sum_{i=1}^{m}\sum_{q=1}^{m}I_{iq}^C(C_{iq}-g(P_i^T X_q))^2 \\
&+\frac{\lambda_S}{2}\sum_{j=1}^{n}\sum_{k=1}^{n}I_{jk}^S(S_{jk}-g(E_j^T Z_k))^2+\frac{\lambda_P}{2}\|P\|_F^2 \\
&+\frac{\lambda_E}{2}\|E\|_F^2+\frac{\lambda_X}{2}\|X\|_F^2+\frac{\lambda_Z}{2}\|Z\|_F^2, \quad (19)
\end{aligned}$$

where $\lambda_C = \sigma_R^2/\sigma_C^2$, $\lambda_S = \sigma_R^2/\sigma_S^2$, $\lambda_P = \sigma_R^2/\sigma_P^2$, $\lambda_X = \sigma_R^2/\sigma_X^2$, $\lambda_E = \sigma_R^2/\sigma_E^2$, and $\lambda_Z = \sigma_R^2/\sigma_Z^2$. A local minimum of the objective function given by Eq. 19 can be found by

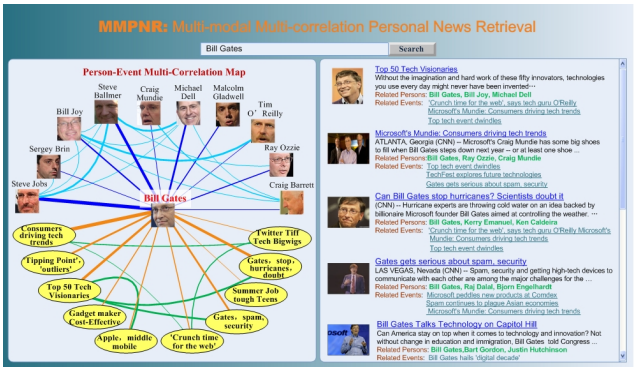


**Figure 6: Interface of MMPNR.**

performing gradient descent in $P_i$, $E_j$, $X_q$ and $Z_k$, respectively.

$$\frac{\partial L}{\partial P_i} = \sum_{j=1}^{n} I_{ij}^{R} g'(P_i^T E_j)(g(P_i^T E_j) - R_{ij}) E_j + \lambda_C \sum_{q=1}^{m} I_{iq}^{C} g'(P_i^T X_q)(g(P_i^T X_q) - C_{iq}) X_q + \lambda_P P_i$$

$$\frac{\partial L}{\partial E_j} = \sum_{i=1}^{m} I_{ij}^{R} g'(P_i^T E_j)(g(P_i^T E_j) - R_{ij}) P_i + \lambda_S \sum_{k=1}^{n} I_{ik}^{S} g'(E_j^T Z_k)(g(E_j^T Z_k) - S_{jk}) Z_k + \lambda_E E_j$$

$$\frac{\partial L}{\partial X_q} = \lambda_C \sum_{i=1}^{m} I_{iq}^{C} g'(P_i^T X_q)(g(P_i^T X_q) - C_{iq}) P_i + \lambda_X X_q$$

$$\frac{\partial L}{\partial Z_k} = \lambda_S \sum_{j=1}^{n} I_{jk}^{S} g'(E_j^T Z_k)(g(E_j^T Z_k) - S_{jk}) E_j + \lambda_Z Z_k$$

To reduce the model complexity, in all of our experiments, we set $\lambda_P = \lambda_E = \lambda_X = \lambda_Z$.

# 6. RANKING AND VISUALIZATION

Provided with the latent feature spaces by MPMF, we can give users more information than the traditional news engines, which can only present the list of related news articles. Figure 6 gives the interface of our MMPNR system. It presents the discovered relations and the relative news events to the end users in a visualized view. Basically, it comprises two types of views: relation view and relative event view.

In the relation view, we give users three relations to answer their queries: person relation, event relation and query person-event relation. In the person relation part, we present a social network about the most relevant persons, which enables users to explore highly relevant information during searching to discover interesting relationships about persons associated with their queries. We also show users a news event relation map about the most related events in the event relation part. Through MPMF, we have got the latent spaces $P$, $E$, $X$ and $Z$, which can be utilized to reconstruct the three correlation matrices by the following formulas:

$$\hat{R} = g(P^T E) \qquad \hat{C} = g(P^T X) \qquad \hat{S} = g(E^T Z).$$

If user submits a query corresponding to the person $i$ in our dataset, we can rank persons and events by sorting the $i$-th column of $\hat{C}$ and $\hat{R}$ by descending order. We can also derive the relevant events to the query from the matrix $\hat{S}$. We only present the top 10 relevant persons in the social network and the top 10 relevant events in the news event relation map. As shown in the left part of Figure 6, we give names and face images of persons and keywords of events. The weighted edges between persons or events denote the relations between them. The thicker the line between persons or events, the stronger the relation they have. The query person-event view shows the relations between the relative news events and the query person using weighted edges. Users can also see the detailed information about a specific event or person through putting the mouse pointer on the suitable position.

In the relative events view, as done in a traditional news searcher, we also present a ranking list of relative news events with general introduction. We present news event not only with the title and a shot part of summary similar to the traditional news searcher, but also with the top 3 relevant persons and the top 3 relevant news events, which can be obtained by sorting the reconstructed event correlation matrix $\hat{S}$. Users can browse more information through clicking the title of events.

# 7. EXPERIMENTS

The objective of our experiments is to examine the effectiveness of our proposed models in person-centric news retrieval. we first explain the data set we collected for our evaluation, the metrics, and the parameter setting in our experiments. Then we present the experimental results using our algorithm as well as its comparison with other methods.

## 7.1 Experimental Design

Our experiments are performed on a web news dataset, in which news articles were crawled from ABCNews.com, BBC.co.uk and CNN.com. Two news articles are considered to be duplicate when they correspond to the same news event according to the '4W' criterion. With the crawled news articles, we first remove the duplicate ones and conduct the evaluation on the deduplicated dataset. That is, one news article stands for a news event in our experiments. We got 99,885 articles in total, whose distribution over the three websits is shown in Table 1. In addition, we extracted 9,345 persons' names as the entities from the news dataset after deleting the ones which appear less than 10 times.

**Table 1: Details of our web news dataset**

| Web site | ABC | BBC | CNN | Total |
|---|---|---|---|---|
| Number of articles | 47,163 | 11,073 | 41,649 | 99,885 |

Similar to previous work on information retrieval, we adopt normalized Discounted Cumulative Gain (nDCG) [13] as a measure to evaluate the effectiveness of web search algorithm, which is defined as

$$\text{nDCG@}k = \frac{\text{DCG}[k]}{\text{IDCG}[k]}.$$

DCG (Discounted Cumulative Gain) is to measure the cumulative gain of the resulting documents on its position and IDCG is the ideal discounted cumulative gain vector. At last, nDCG values for all queries can be averaged to obtain a measure of the average performance of a ranking algorithm.

No well-defined ground-truth dataset can be used to evaluate the performance of news retrieval. Thus, we invite a group of ten people to judge the relevance of searching results. As defined in Table 2, the participators can present three types of graded relevance, which is used in the calculation of nDCG. The specific task of each participator is to randomly select ten queries from the query list as shown in Table 3 to search news information and evaluate the performance according to our predefined evaluating criterions.

There are some parameters to be set in advance. We set $\alpha = 0.5$, $\beta = 0.3$, $\gamma = 0.4$, $\lambda_P = \lambda_E = \lambda_X = \lambda_Z = 0.001$, $\lambda_C = 10$, $\lambda_S = 25$ and $d = 100$. The initial values for $P$ are set by Random Acol [14] through averaging 1,000 columns randomly chosen from $R$. The matrices $E$, $X$ and $Z$ are initialized similarly.

**Table 2: The Graded Relevance**

| *Relevance Level* | *Weight* |
|---|---|
| Very relative | 3 |
| Relative | 2 |
| Irrelative | 1 |

## 7.2 Comparison on Retrieval Performance

We perform the experimental comparisons among six retrieval systems. They are MMPNR considering both person correlation and event correlation (our proposed system), MMPNR-Text only employing text information in correlation initialization, EPMF only considering event correlation, PPMF only considering person correlation, and PMF with no consideration of the both correlations. Additionally, Google News search engine[5] is employed as a baseline in the comparison.

In the first experiment, the participators were asked to search on the six systems to give scores about the relevance to queries as shown in Table 2. The ranking quality is measured using the average nDCG@$k$ for $k$ from 1 to 10. Figure 7 presents the average scores on the top 10 events returned for each query. Figure 8 presents the corresponding gains of ranking quality over baseline. From Figure 7 and Figure 8, we can draw the following observations. First, all the factor analysis based methods achieve the superior effectiveness over Google News, which only considers the text relevance to a given query. Among these, the proposed MMPNR achieves the best performance by simultaneously employing the three correlations and multi-modal information analysis. Second, the worse performance is achieved by MMPNR-Text compared with MMPNR. This demonstrates that the process of face detection and matching is useful in the entity correlation initialization. Third, with additional consideration of entity (or event) correlation, PPMF (or EPMF) obtains more attractive performance than PMF. Fourth, PMF is better than Google news when $K$ is no less than 4, because PMF is able to find more relevant results by mining hidden correlations between entities and events. However, when $K$ is less than 4, i.e., the top 1-3 returned results is considered, a little worse performance is obtain by PMF compared with Google News. It is understandable because Google News adopt a more comprehensive ranking strategy (e.g., PageRank and log analysis in search) to measure the query-oriented relevance.

[5]http://news.google.com/

**Table 3: The query list used in experiments**

| | | |
|---|---|---|
| Allen Craig | Andre Owens | Ayrton Senna |
| Barack Obama | Blake Griffin | Bobby Simmons |
| Caster Semenya | Charlie Villanueva | Chase Utley |
| Chelsea Clinton | Chris Samuels | Christopher Dodd |
| Christopher Plummer | Clarence Thomas | Claudio Pizarro |
| Cole Aldrich | Darren Fletcher | David Brinkley |
| Eddie Griffin | Edgar Davids | Edison Miranda |
| Elie Wiesel | Emily Blunt | Eric Schmidt |
| Frank Lloyd | Gene Hackman | George Washington |
| Greg Kinnear | Harry Hopkins | Howard Baker |
| Hugh Grant | Imam Khomeini | Indiana Jones |
| Jack Coleman | Jackie Chan | James Baker |
| James Steinberg | Jarno Trulli | Jason Kendall |
| Jerry Siegel | Jesse Ventura | John Conyers |
| John Huston | John McGraw | John Paul Jones |
| Julie Christie | Justin Timberlake | Katie Hoff |
| Kelly Clarkson | Landon Donovan | Larry McReynolds |
| Lionel Messi | Lord Mandelson | Mark Hatfield |
| Martin Demichelis | Meredith Whitney | Mike Schmidt |
| Neil Armstrong | Ottoman Sultan | Patrick Cowan |
| Penelope Cruz | Peter Bergen | Prince William |
| Randi Weingarten | Robert Kubica | Samantha Ronson |
| Sarah Ferguson | South Vietnam | Stephen Hendry |
| Steve Cohen | Steve Jobs | Terry Nichols |
| Troy Murphy | Wayne Rooney | William Wallace |

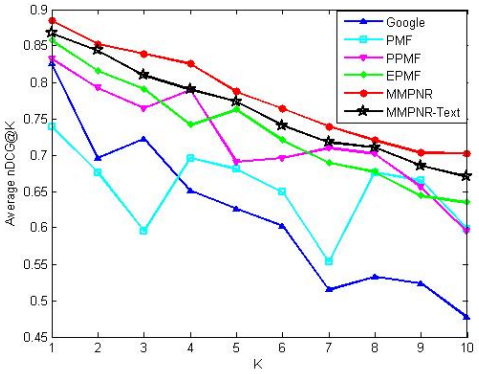


**Figure 7: Comparison on nDCG@k**

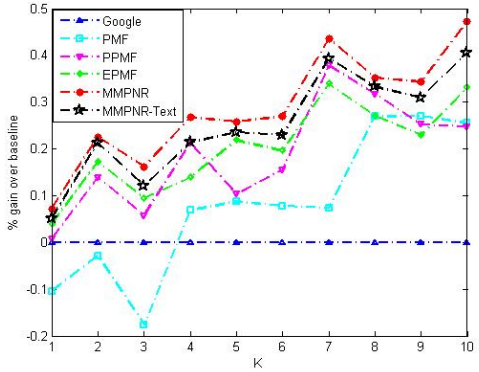


**Figure 8: Gains over Google News.**

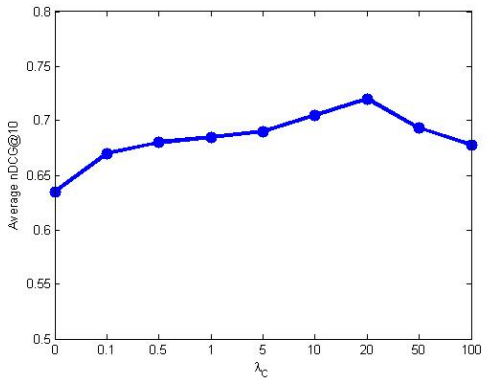


**Figure 9: Impact of Parameter $\lambda_C$ on MMPNR**

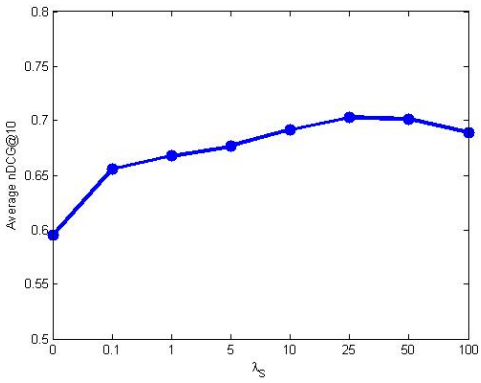


**Figure 10: Impact of Parameter $\lambda_S$ on MMPNR**

Second, we evaluate our system from a subjective view. While browsing results of the ten queries, each participator was asked to give a score within the interval [1, 5] (the bigger the score, the better the result) based on the following aspects:

- *Relevancy*: How about the whole relevancy between the results and queries? Are the results useful?
- *Person Relevancy*: How about the social network? Are the presented persons relative to the query person?
- *Event Relevancy*: How about the event relation map? How about the relevance among the presented events?
- *Efficiency*: How long does the system cost for each query?
- *Friendliness*: Do the users enjoy the interface? Does the interface seem comfortable?
- *Convenience*: Is it convenient to search and browse the news?
- *Multiplicity*: Does the system show users many kinds of information? Can it present users multi-view effectively?

The average scores are shown in Figure 11. It is obvious that users prefer the interface of our system and our system can give relevant and various results conveniently and efficiently. Additionally, the users are satisfied with the presented social network and the relevant event map, which demonstrates the MPMF model is effective to mine the hidden relationships. Additionally, our system is useful and convenient for user to understand news events.

In summary, the above experiments demonstrate that our system is able to mine more relations and give users multi-view relations. They can more easily understand the news events and obtain more information about their queries.

## 7.3 Discussion on Parameters of $\lambda_C$ and $\lambda_S$

The main advantage of our person-centric news retrieval is that it incorporates the social network information and event correlation information. In our model, parameters $\lambda_C$ and $\lambda_S$ balance the information from the person-event relation matrix, person social network and event correlation matrix. If $\lambda_C = 0$ (or $\lambda_S = 0$), it is equivalent to EPMF

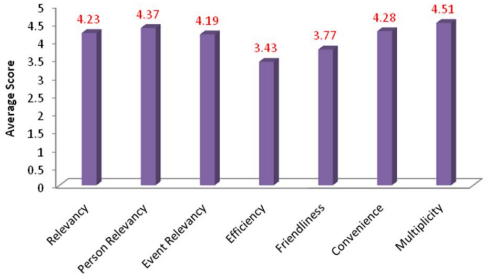


**Figure 11: The average scores of the subjective evaluation**

(or PPMF), respectively. Figure 9 and Figure 10 show the impacts of $\lambda_C$ (while holding $\lambda_S = 25$ fixed) and $\lambda_S$ (while holding $\lambda_C = 10$ fixed) on nDCG@10 respectively, which demonstrate that the values of $\lambda_C$ and $\lambda_S$ impact the results. As $\lambda_C$ increases, the average nDCG@10 increases at first, but when $\lambda_C$ surpasses a certain threshold, the average nDCG@10 decreases with further increase of the value of $\lambda_C$. The impact of parameter $\lambda_S$ is similar to $\lambda_C$. This phenomenon coincides with the intuition that purely using the person-event relation matrix, the person social network or the event correlation matrix can not generate better performance that fusing these three sources.

# 8. CONCLUSIONS

In this paper, we propose a news retrieval approach based on multi-modal analysis and multi-correlation exploration. We explore the information of both text and corresponding images for the initialization of entity correlation. In particular, we adopt statistical co-occurrences in the two modalities to derive the news entity correlations. To fully employ the multi-correlation information, we proposed a Multi-correlation Probabilistic Matrix Factorization model (MPMF) to analyze news entity correlation, news event correlation and entity-event correlation. The proposed MPMF model explores the three correlations simultaneously to rank news events and news persons relevant to a given query. We build a news retrieval system named MMPNR (Multi-modal Multi-correlation Person-centric News Retrieval) based on it. The reasonable and comprehensive evaluations are performed to demonstrate the effectiveness of our system.

The MPMF model opens a broad way for future improvement and extension. As part of future work, we will investigate the following directions: (i) kernel based representation for the two low-dimensional vectors; (ii) considering the information diffusion between persons (events). We believe they would lead us to more promising results.

## 9. ACKNOWLEDGE

This work was supported by the National Natural Science Foundation of China (Grant No. 60903146, 60723005 and 90920303), and 973 Program (Project No. 2010CB327905).

## 10. REFERENCES

[1] E. Agichtein and L. Gravano. *Snowball*: Extracting relations form large plain-text collections. *DL 2000*, pages 85–94, 2000.

[2] M. Banko, M. Cafarella, S. Soderland, M. Broadhead, and O. Etzioni. Open information extraction from the web. *IJCAI*, pages 2670–2676, 2007.

[3] M. Banko and O. Etzioni. The tradeoffs between open and traditional relation extraction. *ACL*, pages 28–36, 2008.

[4] S. Bin. Extracting patterns and relations from the world wide web. *International Workshop on the Web and Databases*, pages 172–183, March 1998.

[5] Y. Chi, S. Zhu, Y. Gong, and H. Zhang. Probabilistic polyadic factorization and its application to personalized recommendation. *CIKM*, October 2008.

[6] D. Dueck and B. Frey. Probabilistic sparse matrix factorization. In *Technical Report PSI TR 2004-023*. Dept. of Computer Science, University of Toronto, 2004.

[7] O. Etzioni, M. J. Cafarella, D. Downey, S. Kok, A.-M. Popescu, T. Shaked, S. Soderland, D. S. Weld, and A. Yates. Web-scale information extraction in knowitall: (preliminary results). *WWW 2004*, pages 100–110, 2004.

[8] O. Etzioni, M. J. Cafarella, D. Downey, A.-M. Popescu, T. Shaked, S. Soderland, D. S. Weld, and A. Yates. Unsupervised named-entity extraction from the web: An experimental study. *Artificial Intelligence*, 165(1):91–134, 2005.

[9] N. Friedman, L. Getoor, D. Koller, , and A. Pfeffer. Learning probabilistic relational models. *IJCAI*, pages 1300–1309, 1999.

[10] C. Giuliano, A. Lavelli, , and L. Romano. Exploiting shallow linguistic information for relation extraction from biomedical literature. *EACL*, 2006.

[11] A. Harabagiu, C. A. Bejan, , and P. Morarescu. Shallow semantics for relation extraction. *IJCAI*, pages 1061–1067, 2005.

[12] T. Hofmann. Probabilistic latent semantic analysis. *Proceeding of the 15th Conference on Uncertainty in AI*, pages 289–296, 1999.

[13] K. Jarvelin and J. Kekalainen. Cumulated gain-based evaluation of ir techniques. *ACM Transactions on Information Systems*, 20(4):422–446, 2002.

[14] A. N. Langville. Algorithms for the nonnegative matrix factorization in text mining. *Slides from SAS Meeting*, 2005.

[15] Z. Li, B. Wang, M. Li, , and W.-Y. Ma. A probabilistic model for retrospective news event detection. *SIGIR*, pages 106–113, 2005.

[16] Y.-R. Lin, J. Sun, P. Castro, R. Konuru, H. Sundaram, , and A. Kelliher. Metafac: community discovery via relational hypergraph factorization. *SIGKDD*, pages 527–536, 2009.

[17] C. Liu, J. Yuen, A. Torralba, , and J. Sivic. Sift flow: dense correspondence across different scenes. *ECCV*, October 2008.

[18] B. Long, Z. Zhang, , and P. Yu. A probabilistic framework for relational clustering. *SIGKDD*, pages 470–479, 2007.

[19] H. Ma, M. R. L. H. Yang, and I. King. Sorec: social recommendation using probabilistic matrix factorization. *CIKM 2008*, pages 931–940, 2008.

[20] B. Marlin. Modeling user rating profiles for collaborative filtering. *NIPS*, 2003.

[21] B. Marlin and R. S. Zemel. The multiple multiplicative factor model for collaborative filtering. *ICML 2004*, 20(4):422–446, July 2004.

[22] M. Naughton, N. Kushmerick, and J. Carthy. Clustering sentences for discovering events in news articles. *ECIR*, 3936:535–538, 2006.

[23] R. Salakhutdinov and A. Mnih. Probabilistic matrix factorization. *Advances in Neural Information Processing Systems*, 20:4, 2008.

[24] H. Sayyadi, A. Sahraei, and H. Abolhassani. Event detection from news articles. *CSICC 2008*, pages 981–984, 2008.

[25] S. Sekine, K. Sudo, and C. Nobata. Extended named entity hierarchy. *LREC*, pages 1818–1824, 2002.

[26] S. Soderland, O. Etzioni, T. Shaked, , and D. Weld. The use of web-based statistics to validate information extraction. *ATEM 2004*, 2004.

[27] C. Spearman. "general intelligence", objectively determined and measured. *The American Journal of Psychology*, 15(2):201–292, 1904.

[28] M. E. Tipping and C. M. Bishop. Probabilistic principal component analysis. In *Technical Report NCRG/97/010*. Neural Computing Research Group, Aston University, September 1997.

[29] Y. Yang, T. Pierce, , and J. G. Carbonell. A study on retrospective and on-line event detection. *SIGIR*, pages 28–36, 1998.

[30] D. Zelenko, C. AoneE, , and A. Richardella. Kernel methods for relation extractions. *Machine Learning Research*, pages 1083–1106, 2003.

[31] Y. Zhang, J. Callan, and T. Minka. Novelty and redundancy detection in adaptive filtering. *SIGIR*, pages 81–88, 2002.

[32] G. Zhou, M. Zhang, D. H. Ji, and Q. Zhu. Tree kernel-based relation extraction with context-sensitive structured parse tree information. *EMNLP-CoNLL*, pages 728–736, 2005.

[33] J. Zhu, Z. Nie, X. Liu, B. Zhang, , and J.-R. Wen. *StatSnowball*: a statistical approach to extracting entity relationships. *WWW 2009*, pages 101–110, 2009.

[34] S. Zhu, K. Yu, Y. Chi, , and Y. Gong. Combining content and link for classification using matrix factorization. *SIGIR 2007*, pages 487–494, 2007.