Applied Mathematics & Information Sciences
An International Journal

http://dx.doi.org/10.12785/amis/092L19

A Personalized News Recommendation using User Location and News Contents

Hee-Geun Yoon[1], *Hyun-Je Song*[1], *Seong-Bae Park*[1,*] *and Kweon Yang Kim*[2]

[1] School of Computer Science and Engineering, Kyungpook National University, Daegu 702-701, Korea
[2] Department of Computer Engineering, Kyungil University, Gyeongsan 712-701, Korea

Received: 18 May 2014, Revised: 19 Jul. 2014, Accepted: 21 Jul. 2014
Published online: 1 Apr. 2015

Abstract: Personalized news article recommendation provides interesting articles to a specific user based on the preference of the user. With increasing use of hand-held devices, the interests of users are not influenced only by news contents, but also by their location. Therefore, in this paper, we propose a novel topic model to incorporate user location into a user preference for the location-based personalized news recommendation. The proposed model is Spatial Topical Preference Model (STPM). By representing the preference of a user differently according to the location of the user, the model recommends the user appropriate news articles to the user location. For this purpose, we represent geographical topic patterns with a Gaussian distribution. STPM is trained only with the news articles that the user actually reads. As a result, it shows poor performance, when the user reads just a few news articles. This problem of STPM is compensated for by LDA-based user profile that is not affected by user location. Therefore, the final proposed model is a combined model of STPM and LDA. In the evaluation of the proposed model, it is shown that STPM reflects user locations into news article recommendation well, and the combined model outperforms both STPM and LDA. These experimental results prove that the location-based user preference improves the performance of news article recommendation, and the proposed model incorporates the locational information of users into news recommendation effectively.

Keywords: Personalized news recommendation, Spatial topical model, Latent Dirichlet Analysis

1 Introduction

These days news reading environment has changed greatly. Web-based news reading services like Google News and Yahoo! News have become increasingly prevalent, as the Internet allows fast access to news articles. Since news articles are in flood from a number of news publishers, news article recommendation has been studied to provide interesting news for readers. However, the interests of readers are different one another. Thus, there have been a number of studies on personalized news article recommendation that focuses more on different interests of readers in recommending news articles [1].

Most personalized news article recommender systems first build user preferences for their service, and then find news articles that are fitted well to the preferences. The systems based on collaborative filtering obtain user preferences from a user-news matrix which consists of news consumption patterns of news readers [1, 2]. The preference of a user is extracted by the patterns of other users who have a similar pattern with the target user. On the other hand, content-based news article recommendation systems construct user preferences by analyzing news contents that they read [3]. In building user preferences, word frequency of news contents has been widely used. However, topics are preferred to word frequency recently, since they are good proxies to news articles and user preferences. As a result, most recent studies employ a topic modeling of user preference [4–6]. In these systems, the similarity between a news article and a user preference is computed by calculating the similarity between their latent topics.

Personalized news article recommendations are successful and show good performances [1, 4]. The users of topic-based personalized recommendations have their own topics that explain their interests. Thus, the recommendation systems can select news articles using the topics. However, they neglect the location of a news reader, even though a location is one of most important elements that determine user preference. With the popularization of hand-held devices such as a mobile

* Corresponding author e-mail: sbpark@sejong.knu.ac.kr

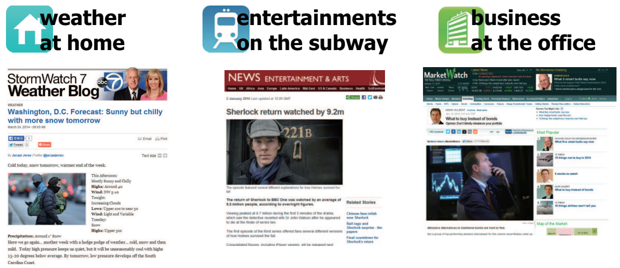


Fig. 1: Different news articles at different user locations.

phone or a tablet PC, people read news articles in which they are interested at any place that they want. In such circumstances, the choice of news articles does not depend only on usual interests of a reader, but also her location. Ordinary people have their own spatial patterns in daily life. For instance, let us consider a person who works for a finance company. She gets up at her home and has brunch at a restaurant near her office; she is working at her office, and often drinks at her favorite bar; lastly, she comes back home to take a rest. Thus, the type of news articles that she reads can be different according to her locations. Figure 1 shows example news articles she read at each location. She often reads entertainments on the subway, and she reads business articles at her office. This is because business articles are related with her work, even if she prefers entertainment news to business news. In addition, she also reads weather news at her home to check the weather of tomorrow. Therefore, it is of importance to consider the location of a news reader as well as the content preference in news article recommendation.

This paper proposes a novel method for topical representation of user preference. The method models geographical patterns of a user jointly with word co-occurrence patterns of news articles. Topic models provide a good description for document contents in general [7]. Thus, the proposed method exploits topic models to build user preference in news recommendation. In addition, the method also employs one other important element that affects user preference. Each article read by a user always has a corresponding location at which the user is positioned. Therefore, topics of a news article should be generated from not only content words but also from a location stamp.

People spend most of their time on a few major spots. Therefore, the locations of a news reader are likely to be somewhere around the spots. To reflect this phenomenon into a topic model, the proposed method parameterizes a Gaussian distribution over locations associated with each topic. In the existing topic modelings, a topic is generated by discovering a word co-occurrence pattern over documents. Since every article has an observable location stamp, each word in the article corresponds to the location stamp of the article. When some words are observed strongly together around a certain spot and do not appear at other places, the topics for them are generated from a Gaussian distribution for the spot. If the spot is a small place, the Gaussian distribution for the spot has a small variance. On the other hand, if the words co-occur at most spots, then the topic for them is generated from a Gaussian distribution with a broad variance. By combining location patterns of a user and word co-occurrence patterns of the news articles, the topics generated by the proposed method become more personalized. That is, the trained topic model reflects the more personalized geographical viewpoint of a user. Therefore, the proposed method recommends news articles in higher quality according to both contextual preference and location of a user. The news articles that a user actually reads are collected with a smart phone. Since every smart phone has a GPS module, a GPS tag is attached to each article. This GPS tag records the location at which the article is read.

One thing to note in personalized news article recommendation is that topics are biased by news articles that a user read. The proposed topic model is trained *only* for each location with the news articles read by the user at the location. Thus, if the user reads just a few news articles at a location, the articles do not provide information enough to discriminate the topics of newly incoming news articles at the location. In order to compensate for this location-based topic model, we employ Latent Dirichlet Allocation (LDA) [8] as another topic representation. LDA is trained using all available news articles. Since a great volume of news articles are available online, it can extract diverse topics from them. That is, the topics from LDA can reflect newly coming articles well. These two topic models of the location-based topic model and LDA recommend new articles by combining their recommendation scores. When a news article is given, each topic model assigns a recommendation score to the article using its own user preference. Then, the final recommendation score of the article is determined by weighted sum of the scores.

A series of experiments were conducted to verify the recommendation performance of the proposed method. In the experiments, we compare the proposed method with its base models. That is, LDA and the location-based topic model are compared with the proposed combined model respectively. A Korean news corpus mined from World Wide Web is used as recommendation candidates. According to our experimental results, the location-based topic model outperforms LDA in topic quality. In addition, it is also shown that the combined method outperforms both LDA and the location-based topic model in recommendation performance. These results prove (i) that the reading preference of a user is influenced by her location and our method reflects this fact effectively, and (ii) that the combination of the two models provides high quality recommendation of news articles.

The rest of this paper is organized as follows. Section 2 reviews the related studies on personalized news article recommendation. Section 3 introduces the problem of location-based personalized news article recommendation. Section 4 and Section 5 explain Spatial Topical Preference Model and Latent Dirichlet Allocation for personalized news are given in Section 6. Finally, Section 7 draws some conclusions.

2 Related Work

Personalized news recommendation has been of interest for a long time. Several adaptive news recommendation services such as Google News and Yahoo! News recommend personalized news articles under the collaborative filtering [1]. Such a success of the collaborative filtering encouraged many studies to propose various enhanced collaborative filtering approaches. Many of them augmented the collaborative filtering in the use of user profiles and news content properties. Liu et al. analyzed user click behaviors and built a user profile on her news interests based on both her past click behavior and the contents that she read [9]. Then, they combined both user click behavior and user profile simultaneously under a Bayesian framework. Their method has been applied to Google News, and the improvement of the recommendation quality was shown over the existing collaborative filtering. Chu et al. tried to analyze dynamic contents at the front page of Yahoo! News according to the passage of time [2]. For this, they constructed the general profile of users using demographic information, activities on relevant sites, and so on. In order to handle the associations between the dynamic contents and the general profile, they proposed a machine learning approach, so-called a feature-based bilinear regression model, and showed that their method outperforms significantly six existing competitive approaches.

Apart from the commercial news recommendation services, a variety of personalized news recommendations have been developed also in academic communities [4, 10]. Li et al. modeled the personalized recommendation of news articles as a contextual-bandit problem [10]. They selected articles sequentially to serve users based on the contextual information, click behavior of users, and click through rates of articles. They verified their method by showing that their method provides personalized web service effectively on traffic of the front page of Yahoo! News. On the other hand, Li et al. proposed two-stage personalized news recommendation systems [4]. In order to deal with large volume of newly published news collections, they employed minhashing and locally sensitive hashing to reduce the number of comparisons in the first stage. In addition, they used a hierarchical clustering for further speeding up of news article selection. In the second stage, they recommended newly published articles by considering not only the contents of news articles but also the entity preferences such as people and events.

Recently many recommendation services try to incorporate a user context into choosing interesting articles. Lin et al. recommended news articles by incorporating three factors into a probabilistic matrix factorization [11]. The three factors considered are news content information, collaborative filtering, and information diffusion in virtual social network. In this work, the information from social networks is regarded as “word of mouth” that can be used to help making a decision for unexperienced users. Li and Li also tried to reflect the user behavior and news content simultaneously for news recommendation [12]. In this study, they employed a unified graph-based approach to model multi-type objects such as event types, topics, and implicit relations in news reading community. They proved through an experiment on a data set collected from various news websites that their method handles the cold-start problem effectively.

Many previous studies have considered not just content preference but also some other factors such as user click behavior, entity preferences, and social relationship of users. These studies have proven the effectiveness of those factors for personalized news recommendation. However, none of them consider the spatial context of users even though users’ interests change according to their location. That is, different user preferences should be discovered with respect to user locations. Therefore, it is important to regard the spatial context as one of contexts for news recommendation. Then, the news recommendation can reflect the routine life of users, and derive high quality of news article recommendation.

3 Personalized News Article Recommendation with User’s Spatial Information

Figure 2 describes the overall process of measuring how well a news article a is fitted to the user preference with her location l. The news articles that a user has already read are assumed to be given in advance, and all the articles are to be geo-tagged. Here, the tag of an article comes from user location at which the user read the article. When the geo-tagged news articles are given, two kinds of topic models predict the relevance score of a new news article a at the current user position l respectively. One is Spatial Topic Preference Model (STPM) that predicts the score with the consideration of the current user location, and the other is Latent Dirichlet Allocation (LDA) that predicts it without location information. The final score of the article a is then determined by weighted sum of their scores.

In STPM, the user preference should be known first. The topical representation of user preference is denoted as $\mathscr{P}^S$ in this figure. $\mathscr{P}^S$ is expressed as a topic vector.

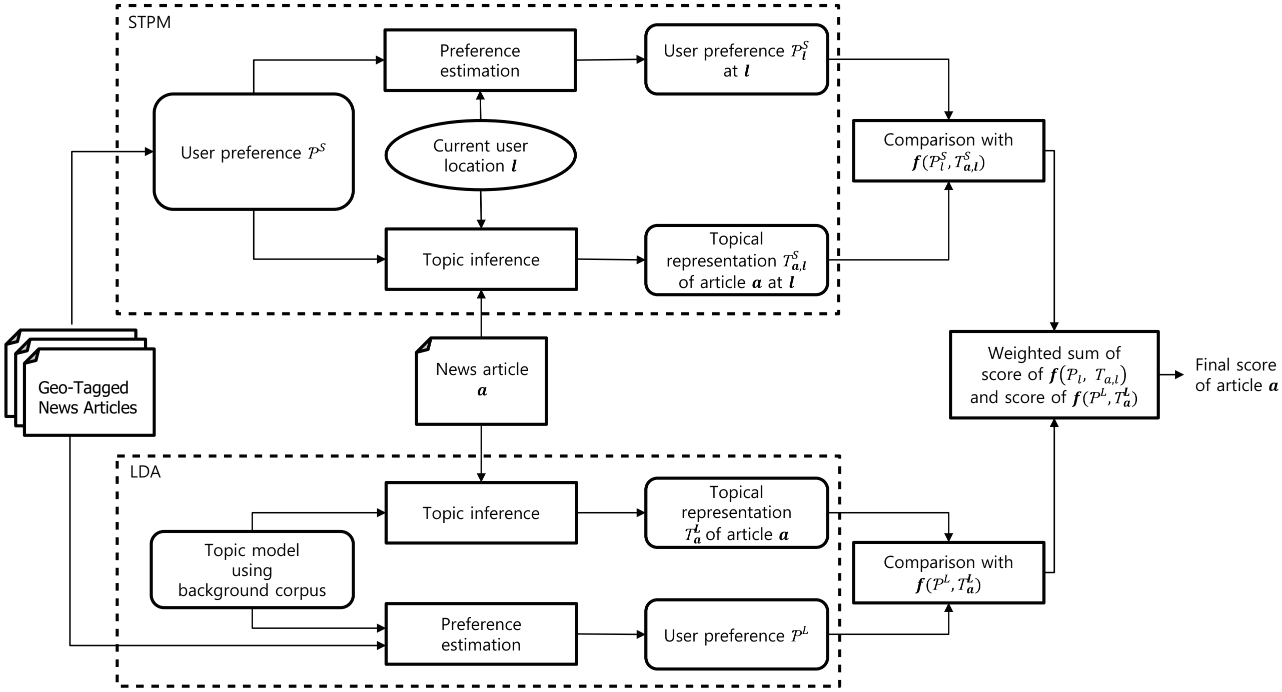


Fig. 2: Overall process of the location-based personalized news articles recommendation

This user preference changes whenever the user moves her locations, so that $\mathscr{P}^S$ should be changed according to user location. That is, when a current user location l is given, the topical representation of user preference at a location l, $\mathscr{P}^S_l$, is estimated from $\mathscr{P}^S$ using locational information. The locational information at the location l is represented as topics of the news articles read at l. Therefore, $\mathscr{P}^S_l$, the user preference at l, is also a topic vector.

An article can be represented as a topic vector. That is, a new article a is also represented as a topic vector $\mathscr{T}^S_a$. The topics of a are affected by the user location l. As a result, we modify it to $\mathscr{T}^S_{a,l}$ by incorporating the location into $\mathscr{T}^S_a$. The appropriateness of $\mathscr{T}^S_{a,l}$ to $\mathscr{P}^S_l$ is determined by a score function $f(\mathscr{P}^S_l, \mathscr{T}^S_{a,l})$. Since both user preference and the representation of a news article are represented as topic vectors, a simple vector similarity is used as the score function. That is, the score function $f(\mathscr{P}^S_l, \mathscr{T}^S_{a,l})$ is computed by

$$f(\mathscr{P}^S_l, \mathscr{T}^S_{a,l}) = \frac{\mathscr{P}^S_l \cdot \mathscr{T}^S_{a,l}}{\parallel \mathscr{P}^S_l \parallel \cdot \parallel \mathscr{T}^S_{a,l} \parallel}. \quad (1)$$

Obviously high-scored articles are more likely to be recommended to the user. In LDA, the general preference of the user is identified. The topical representation of this preference is denoted as $\mathscr{P}^L$ in this figure. Unlike $\mathscr{P}^S_l$, $\mathscr{P}^L$ is not affected by user location. Thus, the topics of *LDA* are estimated using all the news articles available in World Wide Web, and the articles need not be geo-tagged.

Then, $\mathscr{P}^L$, the personalized preference, is generated using the LDA topics and the articles that the user read. However, it is personalized into As in STPM, a new article a is represented as a topic vector $\mathscr{T}^L_a$. After that, the appropriateness of $\mathscr{T}^L_a$ to $\mathscr{P}^L$ is determined by the score function $f(\mathscr{P}^L, \mathscr{T}^L_a)$ given in Equation (1).

The final score of the news article a is computed by combining the scores of STPM and LDA. This score is calculated by their weighted sum. That is, $score(a)$, the final score of a is calculated by

$$score(a) = \gamma \cdot f(\mathscr{P}^S_l, \mathscr{T}^S_{a,l}) + (1-\gamma) \cdot f(\mathscr{P}^L, \mathscr{T}^L_a) \quad (2)$$

where $0 \leq \gamma \leq 1$ is an importance ratio between STPM and LDA. Note that both score functions of STPM and LDA are bounded between 0 and 1 by Equation (1). Thus, the weighted sum is concave and the article with a high final score is more likely to be recommended to the user.

If the words generated by the topics of STPM are similar to those in an article, STPM is believed to express the article well. That is, the larger the number of words that appear both in the topics of STPM and the article is, the more trustworthy STPM is. Therefore, γ can be induced from the overlapped words between STPM and the article a, since it is a parameter to control the importance between STPM and LDA. That is, γ is defined as

$$\gamma = \frac{|\mathbf{w}_m \cap \mathbf{w}_a|}{|\mathbf{w}_a|},$$

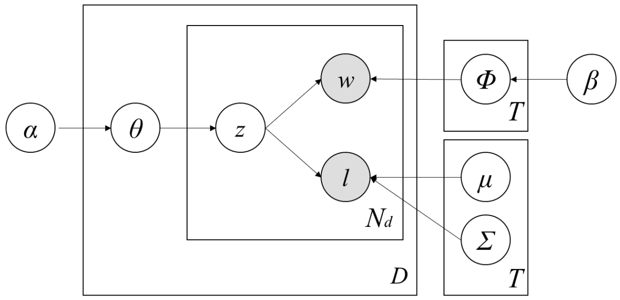


Fig. 3: The graphical model of STPM.

Table 1: Description of the notations used in the graphical representation of STPM.

Symbol	Description
T	the number of topics
D	the number of documents
V	the number of unique words
N_d	the number of words in document d
θ_d	multinomial distribution of topics specific to the document d
Φ_z	multinomial distribution of words specific to topic z
μ_z, Σ_z	Gaussian distribution for a location specific to topic z
z_{di}	topic associated with the i-th word in d
w_{di}	the i-th word in d
l_{di}	location associated with the i-th word in d

where $\mathbf{w}_m$ is a set of words generated from STPM topics, and $\mathbf{w}_a$ is a set of unique words appearing at a.

4 Spatial Topical Preference Model for Personalized News Article Recommendation

Most existing content-based news recommender systems consider only words of articles. Thus, they employ a topic model to find hidden meanings among words. The discovered topics are used to represent a user preference. This is because the topics are generated from word co-occurrence patterns over news articles that were read by the target user. However, topic discovery is not affected only by word co-occurrences. Some previous studies dealt with other factors for topic discovery [13]. In case of constructing the user preference, the user's spatial patterns of reading articles also affect user preference. Thus, STPM reflects the geographical patterns into news recommendation.

Figure 3 shows the graphical model of STPM. The symbols used in this figure are explained in Table 1. Every word is assumed to have its own location stamp. Then, all words are located in a geographical space. However, the number of locations in which a user is moving is not large. Normally a user is active on a few major spots. Therefore, the locations of words appearing in the articles read by the user are somewhere around the several spots. As a result, the topics discovered from the words have geographical patterns. Then, STPM can cluster words through the geographical patterns. That is, the words are clustered by their co-occurrences on their geographical locations as well as the co-occurrences in the articles.

We assume that these geographical patterns of topics are ruled by a Gaussian distribution. The distribution allows the topics of STPM to be flexible, since it is continuous. Note that the location stamps for training STPM are sparse compared to potential locations of a user. By adapting a Gaussian distribution, STPM can avoid the sparsity problem of the locational information. In addition, since each word corresponds to several location stamps, the Gibbs sampling can be used to infer the posterior distribution of the topics.

Let Φ_z be a probability vector of words in a topic z, and $\Phi_{1:T}$ be a probability matrix of T topics and all words. Under STPM, the words and the locations of a document d arise from the following generative process with Dirichlet parameters α and β.

1. For each topic z, draw T word distributions by $\Phi_z \mid \beta \sim Dirichlet(\beta)$.
2. Draw topic proportion $\theta_d \mid \alpha \sim Dirichlet(\alpha)$.
3. For each word w_{di} and its location l_{di} of the document d,
   (a) Draw a topic by $z_{di} \mid \theta_d \sim Multinomial(\theta_d)$.
   (b) Draw a word by $w_{di} \mid z_{di}, \Phi_{1:T} \sim Multinomial(\Phi_{z_{di}})$.
   (c) Draw a location by $l_{di} \mid z_{di}, \mu_{1:T}, \Sigma_{1:T} \sim Normal(\mu_{z_{di}}, \Sigma_{z_{di}})$.

The posterior distribution $P(\theta, \Phi, \mathbf{z}|\mathbf{w}, \mathbf{l})$ should be inferred to train the model, where $\mathbf{w}$ is all words in the articles and $\mathbf{l}$ represents all possible locations. That is, since only words and location stamps are observed in the data set, all other latent variables should be estimated from them. However, the inference cannot be done exactly, since θ, Φ, and $\{\mu, \Sigma\}$ are coupled together in the summation over latent topics. Thus, the Gibbs sampling is used as an alternative method. We sample first the assignments of words and locations to topics, and then θ and Φ are integrated after topics are sampled. Therefore, the conditional posterior distribution for z_{di} is given by

$$\begin{aligned} & P(z_{di}|\mathbf{w}, \mathbf{z}_{-di}, \mathbf{l}) \\ \propto\ & P(w_{di}|\mathbf{w}_{-di}, \mathbf{z}_{-di}) P(l_{di}|\mathbf{l}_{-di}, \mathbf{z}_{-di}) P(\mathbf{z}_{-di}) \\ \propto\ & \frac{n_{z_{di}w_{di}} + \beta_{w_{di}} - 1}{\sum_{v=1}^{V} (n_{z_{di}v} + \beta_v) - 1} (m_{dz_{di}} + \alpha_{z_{di}} - 1) \\ \times\ & \frac{1}{2\pi\sqrt{|\Sigma_{z_{di}}|}} \exp\left(\frac{-(l_{di} - \mu_{z_{di}})^T \Sigma_{z_{di}}^{-1} (l_{di} - \mu_{z_{di}})}{2} \right), \end{aligned} \quad (3)$$

where $\mathbf{z}_{-di}$ is the assignment of all words except w_{di}. In addition, n_{zv} represents the number that a word v is assigned to a topic z, and m_{dz} is the number that the words in an article d are assigned to a topic z. The Gaussian parameters for each topic are obtained by a maximum likelihood method [14].

4.1 User Preference Estimation over User Location

Many previous studies defined the preference of a user as a topic proportion of news articles read by the user [4, 5]. The topic proportion can be represented as a vector whose element is a probability of a topic. This probability is then calculated by summing the conditional posterior distribution for the topic over the news articles.

Formally, a user preference is represented as the topic proportion $\mathscr{P}$, which is expressed as a vector

$$\mathscr{P} = \langle P(z_1), \dots, P(z_i), \dots, P(z_T) \rangle, \quad (4)$$

where T is the number of topics and $P(z_i)$ is the i-th topic probability. When a user reads a set of new articles D, the topic probability $P(z_i)$ is obtained as follows.

$$\begin{aligned} P(z_i) &= \sum_{d \in D} P(z_i|d)P(d) \\ &= \sum_{d \in D} P(z_i|d), \end{aligned} \quad (5)$$

where $P(d)$ is the probability of an article $d \in D$ that is simply assumed to be uniform. $P(z_i|d)$ is usually a conditional posterior distribution like Equation (3).

In STPM, the user preference changes whenever the user moves her locations. Thus, the geographical information should be reflected into user preference. Then, $\mathscr{P}_l^S$, the user preference at a location l is defined as

$$\mathscr{P}_l^S = \langle E(\theta_{z_1}|l), E(\theta_{z_2}|l), \dots, E(\theta_{z_T}|l) \rangle,$$

where $E(\theta_{z_k})$ is an expectation of θ_{z_k}. By the Bayes rule,

$$\begin{aligned} E(\theta_{z_k}|l) &= P(z_k|l) \\ &\propto p(l|z_k)P(z_k). \end{aligned}$$

Here, $p(l|z_k)$ is a conditional probability of l given a topic z_k. Since we assume that geographical information is ruled by a Gaussian distribution, $p(l|z_k)$ becomes

$$p(l|z_k) \propto \frac{1}{2\pi\sqrt{|\Sigma_{z_k}|}} \exp\left(\frac{-(l-\mu_{z_k})^T \Sigma_{z_k}^{-1} (l-\mu_{z_k})}{2}\right).$$

Then, $\mathscr{P}_l^S$ is compared with a topic representation $\mathscr{T}_{a,l}^S$ of an article a by Equation (1). Note that $||\mathscr{P}_l^S||$ is not always one, while $||\mathscr{T}_{a,l}^S|| = 1$. Therefore, to make them be compared, each $E(\theta_{z_k}|l)$ of $\mathscr{P}_l^S$ is normalized to a sum of one. That is, $\mathscr{P}_l^S$ is actually

$$\mathscr{P}_l^S = \left\langle \frac{E(\theta_{z_1}|l)}{Z}, \frac{E(\theta_{z_2}|l)}{Z}, \dots, \frac{E(\theta_{z_T}|l)}{Z} \right\rangle,$$

where $Z = \sum_{i=1}^T E(\theta_{z_i}|l)$.

4.2 Topic Inference of News Articles with User Location

Topic representation $\mathscr{T}_{a,l}^S$ of a news article a with user location l can be obtained by computing a posterior distribution θ of $\{a,l\}$. Hence, we also conduct a sampling procedure for each word in $\{a,l\}$ by Equation (3). Then, $\mathscr{T}_{a,l}^S$ is obtained by marginalizing those sampled topics. Since $\mathscr{T}_{a,l}^S$ is a topic proportion of $\{a,l\}$, it is represented as a vector

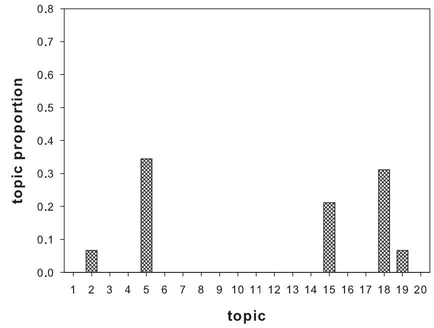
$$\mathscr{T}_{a,l}^S = \langle \theta_{z_1}^{\{a,l\}}, \theta_{z_2}^{\{a,l\}}, \dots, \theta_{z_T}^{\{a,l\}} \rangle.$$

Then, each topic-document distribution $\theta_{z_k}^{\{a,l\}}$ is obtained by counting z_k among sampled topics as follows.

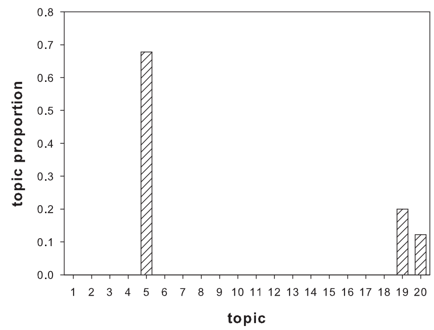
$$\theta_{z_k}^{\{a,l\}} = \frac{m_{az_k} + \delta}{\sum_{k=1}^T m_{az_k} + T\delta},$$

where δ is a smoothing parameter. We use $\delta = 1$ in this paper.

The current user location is used as a location for news recommendation. Therefore, even the same article is represented as different topic vectors according to locations. As a result, a news article at different locations is considered to be different in our model. For instance, Figure 4 shows topic proportions of an article at two different locations. This figure demonstrates the topics of a user who reads Major League Baseball (MLB)-related articles mostly at the location B and reads general articles



(a) Location A



(b) Location B

Fig. 4: Two different topic representations of the same news article by STPM at two different locations.

Table 2: Statistics of news articles read by users.

	Time span	The number of news articles	The number of unique words	The number of tokens
User1	10/25/2013 – 11/11/2013	231	6,106	32,114
User2	06/12/2013 – 11/11/2013	290	7,161	34,595
User3	10/25/2013 – 11/10/2013	307	6,916	35,213
User4	07/01/2013 – 11/10/2013	175	5,045	22,594
Average		250.75	6,307	31,129

at the location A. Thus, when a new article about a MLB game is given, the topic proportion in Figure 4(a) is a little bit different from that in Figure 4(b). Since the topic 5 is a MLB-related topic, it appears strongly at both locations. Note that the topic 18 which is about overall sports is also relatively strong at the location A. However, it does not appear at all at the location B, since the location B is a MLB-specific location. All probabilities for the topic 18 are reflected in the topic 5 at the location B. That is, STPM reflects the difference of locations well into news recommendation, and more sophisticate recommendation gets possible.

5 Latent Dirichlet Allocation for Personalized News Article Recommendation

Since STPM requires not only the contents of news articles but also their GPS tags, the number of news articles that a user actually read is usually small. Note that the user preference $\mathscr{P}^S$ in Figure 2 is generated from the articles that the user read. If the number of the articles is small, each topic from the articles consists of a small number of words, which results in failure of discriminating newly incoming news articles. Suppose that we have two newly-incoming news article a and b. Their topical representations, $\mathscr{T}^S_{a,l}$ and $\mathscr{T}^S_{b,l}$ are made based on the words of the topics of $\mathscr{P}^S$. If the topics of $\mathscr{P}^S$ have just one or two words, then $\mathscr{T}^S_{a,l}$ and $\mathscr{T}^S_{b,l}$ get similar each other even if a and b are completely different. That is, a and b are distinguished by STPM. Therefore, if the number of geo-tagged news articles is small, STPM can not recommend new articles to the user.

Latent Dirichlet Allocation (LDA), one of the widely-used topic models, can compensate for the problem of STPM. Basically it does not require geo-tagged articles. It is trained with any available documents such as webpages crawled from the web. When this huge background corpus is used to generate topics, a variety of topics can be made and the topics consist of a number of words. Thus, LDA can assist STPM even if it does not process any geographical information. As in STPM, a user preference of LDA for personalized news articles recommendation should be constructed. For this, LDA is first trained using a large background corpus. After that, a topic proportion is estimated over the news articles that a user actually read using the trained LDA. Since only the news articles that the user read are used, this topic proportion becomes a personal preference of the user. Thus, this user preference $\mathscr{P}^L$ is made as a vector of topic probabilities by Equation (4), and the probability of a topic is estimated using a conditional posterior distribution given in Equation (5). In the same manner, $\mathscr{T}^L_a$, the topic representation of a newly-incoming news article a is expressed also as a vector of topic probabilities, where the probability of each topic is computed only from the document a.

6 Experiments

6.1 Experimental Settings

For the experiments of location-based personalized news articles recommendation, four users are engaged in reading news articles from June 17, 2013 to November 11, 2013. They were instructed to read news articles using a smart phone with a GPS function. Thus, a GPS coordinate is attached to a news article as soon as the users read the article. Table 2 shows the statistics of news articles read by the four users. Among the four users, User3 read the greatest number of news articles. The user pore over 307 news articles during 18 days, and the articles she read contains 35,213 words and 6,916 unique words. On the other hand, User4 read the least number of news articles. The number of news article read by User4 is only 175, and 5,045 unique words are found in the news articles. Each user read 250.75 news articles on average. When investigating locations at which the news articles read, it is found out that the number of main spatial spots of User2 and User4 is three: home, office, and food court. On the other hand, that of User1 and User3 is just two: home and office.

In addition, two news corpora are collected from Ziny News[1]. One is as a training data for training LDA, and the other is a repository of news articles to be used as candidate recommendations. Table 3 shows a simple statistics of these two news corpora. The news articles for training LDA, so-called background corpus, contain 103,328 news articles from June 3, 2013 to August 17, 2013. The articles have 56,051 unique words and

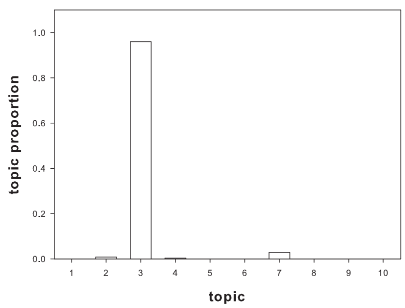
[1] `http://www.zinynews.com`. It is a news reading app. For this service, it collects everyday all news articles that major Korean news agencies publish. Unfortunately, this site provides information on Ziny News only in Korean.

Table 3: Statistics of news article collections for training LDA and recommendation.

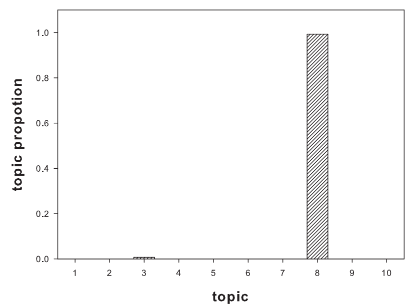
Data set	Time span	The number of articles	The number of unique words	The number of tokens
News articles for LDA	06/03/2013 – 08/17/2013	103,328	56,051	18,215,162
News articles for recommendation	12/01/2013 – 12/31/2013	447	10,597	82,385

Table 4: Top topics of User2 by LDA and STPM

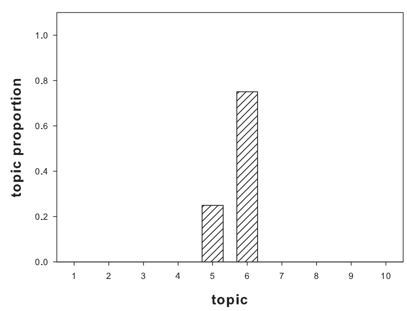
Model	Spot	Topic	Words
STPM	Office	Topic 3 (IT tech)	smart phone, pad, Samsung, virus, apple, ...
	Food court	Topic 8 (hobby)	game, picture, universe, addiction, Japan, ...
	Home	Topic 5 (accident and incident)	police, incident, investigation, syndicate, China, terror, ...
		Topic 6 (politics)	accident, news, America, China, minister, president, ...
LDA		Topic 38 (accident and incident)	police, suspicion, investigation, book, imprison, ...
		Topic 7 (baseball)	Hyun-Jin Ryu, Dodgers, hit, league, pitcher, ...
		Topic 87 (K-pop)	album, music, group, member, singer, star, ...
		Topic 55 (entertainer)	broadcast, picture, confession, woman, reaction, ...



(a) Office



(b) Food court

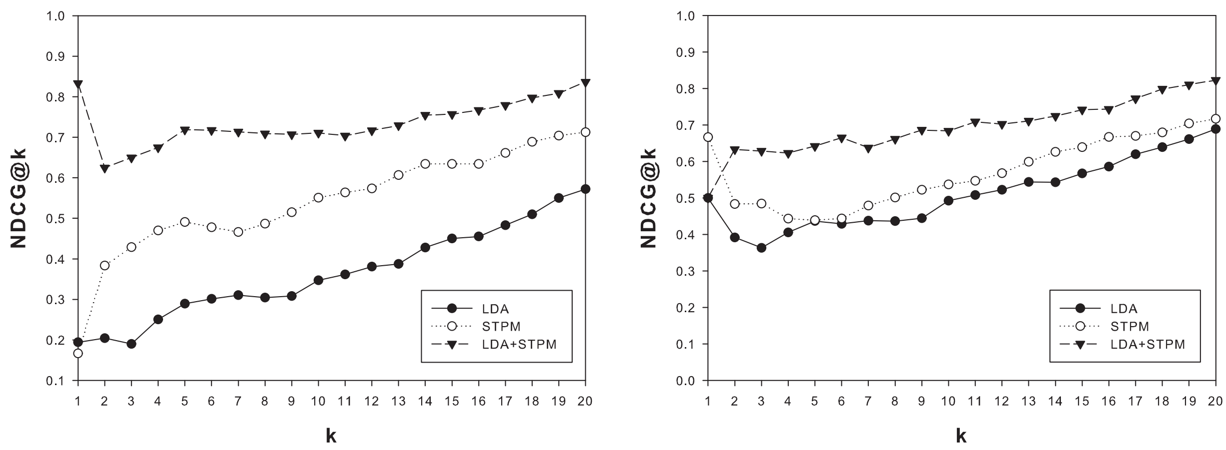


(c) Home

Fig. 5: User preference of User2 at office, food court, and home

18,215,162 tokens in total. Compared to the number of unique words in news articles in Table 2, it is approximately 10 times larger. The recommendation collection consists of 447 news articles that are sampled from the news repository issued from December 1, 2013 to December 31, 2013. This collection has 10,597 unique words and 82,385 tokens.

The superiority of our method is given in two kinds of experiments. First, STPM is compared with LDA to see if STPM reflects user locations into recommendation better than LDA. Since the distinctness of topics represents their quality often [13], the superiority of STPM over LDA is measured by the average distance among discovered topics. That is, topics are more representable and disentangle, if they are more distinct one another. The topic difference is actually measured by average KL-divergence among topics. In the second experiment, in order to see the effectiveness of the proposed method in



(a) Office (b) Home

Fig. 6: Average NDCG@k of four users with various k's at home and at office.

news article recommendation, the proposed combined model is compared with its base models: STPM and LDA. The quality of all models is measured with NDCG@k which is broadly used to evaluate the performance of recommendation systems [15].

In all experiments, the number of topics in LDA is set to be 100. Since the users in Table 2 read just around 250 news articles, STPM can not be trained for more than ten topics. Thus, the number of topics in STPM is fixed to ten. All articles are written in Korean, and only nouns are used for interpreting topics. In addition, for all topic models, a fixed symmetric Dirichlet distribution is used with $\alpha = 1/T$ and $\beta = 0.01$.

6.2 Experimental Results

Table 4 demonstrates the topics of User2 trained by STPM and LDA. It lists top-one topic or top-two topics for each location in STPM, and top-four topics in LDA. Since User2 is a CS-major graduate student, IT-related topic is dominant at office. On the other hand, the topic about hobby is dominant at the food court, and those about society and politics are dominant at home. Figure 5 shows the preference of User2 at different locations. The topic distribution gets definitely different according to the location of the user. Therefore, the preference of STPM represents the user well at each location. On the other hand, the topics generated by LDA are different from those by STPM. A topic, 'accident and incident' is shared by both STPM and LDA, but some topics in STMP such as 'IT tech' and 'politics' do not appear in LDA.

The quality of topics is measured numerically with KL-divergence. KL-divergence measures the average distance of word distributions for all pairs of topics. Thus, the larger KL-divergence between two topics is, the more distinct they are. In our experiments, the average KL-divergence of STPM is 10.90, while that of LDA is 15.70. Since LDA has 100 topics and STPM has just 10 topics, it is natural that the average KL-divergence of LDA is larger than that of STPM. In order to compare them *fairly*, we select 10 topics of LDA that are similar to the topics of STPM. Then, the average KL-diverge is calculated again. The average KL-divergence over the 10 topics of LDA becomes 10.67, which is smaller than that of STPM. Therefore, we can conclude that STPM learns more distinct topics, which implies higher quality of topics.

Note that the final proposed model is the combined one of STPM and LDA. Figure 6 shows the recommendation performance of STPM, LDA, and the proposed model (LDA+STPM). The experiments are conducted with two locations of home and office, and the performance is measured with NDCG@k with various k's. The reason why only two locations are used instead of three locations is that they are the only locations at which all four users read news articles. At 'food court', only two users out of four read news articles. In both Figure 6(a) and 6(b), STPM outperforms LDA. STPM achieves average NDCG@k of 0.57 at office and 0.54 at home. However, LDA scores 0.51 and 0.36 respectively. Since STPM reflects user locations into user preference and topic representation of a news article, it outperforms LDA in both locations. The largest difference between STPM and LDA with respect to NDCG@k is 0.24 at office and 0.17 at home. That is, the difference at office is larger than that at home. This is because there exists only one dominant topic at office, while there are two

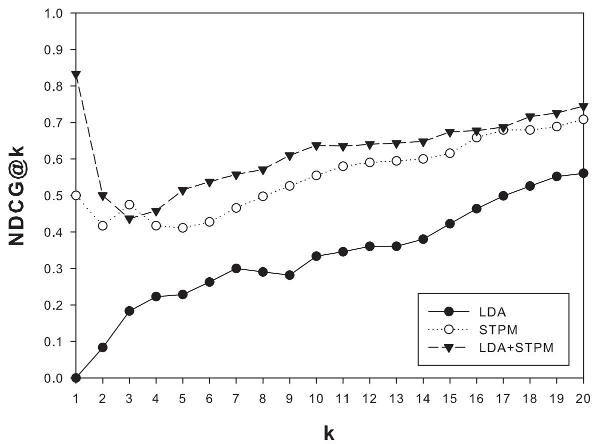


Fig. 7: Average NDCG@k of two users at food court.

dominant topics at home (see Figure 5). On the other hand, the proposed combined model (LDA+STPM) achieves the highest performance for all k's in both locations except $k = 1$ at home. Its average NDCG@k is 0.69 at office and 0.74 at home. This is 0.12 higher than that of STPM and 0.18 higher than that of LDA at office, and 0.19 higher than that of STPM and 0.37 higher than that of LDA at home. The fact that the proposed model outperforms STPM implies that LDA compensates for the problem of STPM well.

Figure 7 depicts average NDCG@k at food court. All models in this figure are trained with the news articles read by two users rather than four users, since only two users read news articles at food court. Even in this figure, the proposed combined model shows the best performance, while LDA achieves the worst performance. Therefore, it can be inferred that the proposed model recommends news articles that fit to both user interests and location.

7 Conclusion

In this paper, we have proposed a novel method for personalized news article recommendation based on current user location. Since spatial information of a user is valuable for constructing preference of the user, we first have proposed the Spatial Topical Preference Model (STPM) that models the spatial information of the user jointly with the word patterns appearing at the news articles read by the user. As a result, STPM generates different user preferences for different locations, and this location-dependent user preference enables STPM to recommend the news articles that are more appropriate to user location.

SPTM is trained only with the news articles read by the user. Thus, it fails in discriminating newly incoming news articles, when the number of the news articles the user read is small. In order to compensate for this problem of STPM, Latent Dirichlet Allocation (LDA) has been adopted. Since LDA is trained with huge background corpus, it can represent news article with a number of words generated from diverse topics. The final recommendation model is obtained by combining STPM and LDA. Therefore, the model can recommend news articles that reflect both user interests and user location, even when the user provides only a few geo-tagged articles.

Through a series of experiments, we have shown that STPM constructs the preferences that are more specialized to user's locational context, and the quality of its preferences is higher than that of LDA. This implies that the preference on news articles of a user is actually influenced by the location of the user, and STPM reflects this fact effectively. In addition, we also showed that the combined proposed model outperforms both base models of LDA and STPM in recommendation performance. These results prove that the simultaneous use of both models enhances the quality recommendation of news articles.

Acknowledgment

This work was supported by the ICT R&D program of MSIP/IITP. [10035348, Development of a Cognitive Planning and Learning Model for Mobile Platforms], [10044494, WiseKB: Big data based self-evolving knowledge base and reasoning platform]

References

[1] A. S. Das, M. Datar, A. Garg, and S. Rajaram, "Google news personalization: Scalable online collaborative filtering," in *Proceedings of the 16th International Conference on World Wide Web*, pp. 271–280, 2007.

[2] W. Chu and S.-T. Park, "Personalized recommendation on dynamic content using predictive bilinear models," in *Proceedings of the 18th International Conference on World Wide Web*, pp. 691–700, 2009.

[3] E. Gabrilovich, S. Dumais, and E. Horvitz, "Newsjunkie: Providing personalized newsfeeds via analysis of information novelty," in *Proceedings of the 13th International Conference on World Wide Web*, pp. 482–490, 2004.

[4] L. Li, D. Wang, T. Li, D. Knox, and B. Padmanabhan, "Scene: A scalable two-stage personalized news recommendation system," in *Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 125–134, 2011.

[5] L. Li, L. Zheng, and T. Li, "Logo: A long-short user interest integration in personalized news recommendation," in *Proceedings of the Fifth ACM Conference on Recommender Systems*, pp. 317–320, 2011.

[6] M. Ovsjanikov and Y. Chen, "Topic modeling for personalized recommendation of volatile items," in *Proceedings of the 2010 European Conference on Machine Learning and Knowledge Discovery in Databases: Part II*, pp. 483–498, 2010.

[7] X. Wei and W. B. Croft, "Lda-based document models for ad-hoc retrieval," in *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 178–185, 2006.

[8] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *Journal of Machine Learning Research*, vol. 3, pp. 993–1022, 2003.

[9] J. Liu, P. Dolan, and E. R. Pedersen, "Personalized news recommendation based on click behavior," in *Proceedings of the 15th International Conference on Intelligent User Interfaces*, pp. 31–40, 2010.

[10] L. Li, W. Chu, J. Langford, and R. E. Schapire, "A contextual-bandit approach to personalized news article recommendation," in *Proceedings of the 19th International Conference on World Wide Web*, pp. 661–670, 2010.

[11] C. Lin, R. Xie, L. Li, Z. Huang, and T. Li, "Premise: Personalized news recommendation via implicit social experts," in *Proceedings of the 21st ACM International Conference on Information and Knowledge Management*, pp. 1607–1611, 2012.

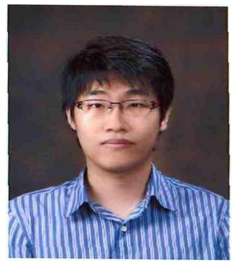
[12] L. Li and T. Li, "News recommendation via hypergraph learning: Encapsulation of user behavior and news content," in *Proceedings of the Sixth ACM International Conference on Web Search and Data Mining*, pp. 305–314, 2013.

[13] X. Wang and A. McCallum, "Topics over time: A non-markov continuous-time model of topical trends," in *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 424–433, 2006.

[14] A. Rencher, *Methods of Multivariate Analysis*. Wiley, 1995.

[15] K. Järvelin and J. Kekäläinen, "Cumulated gain-based evaluation of ir techniques," *ACM Transactions on Information Systems*, vol. 20, no. 4, pp. 422–446, 2002.

Hee-Geun Yoon received his MS in computer engineering from Kyungpook National University, Daegu, Rep. of Korea, in 2009. Now, he is a PhD candidate at the School of Computer Science and Engineering, Kyungpook National University. His main research interests include machine learning and natural language processing.



Hyun-Je Song received his B.S. and MS degree in Computer Engineering from Kyungpook National University in 2008 and 2010 respectively. Currently he is a PhD candidate in School of Computer Science and Engineering, Kyungpook National University. His research interests include machine learning and natural language processing.

Seong-Bae Park received his MS and PhD degrees in computer science from Seoul National University, Seoul, Rep. of Korea, in 1996 and 2002, respectively. Now, he is an associate professor at the School of Computer Science and Engineering, Kyungpook National University, Daegu, Rep. of Korea. He focuses on machine learning, natural language processing, text mining, and bio-informatics.

Kweon Yang Kim received his MS and PhD degrees in Computer Engineering from Kyungpook National University, Daegu, Rep. of Korea, in 1990 and 1998, respectively. Now, he is a professor at the Department of Computer Engineering, Kyungil University, Gyeongsan, Rep. of Korea. He focuses on semantic web and Korean language engineering.