Location-Aware Personalized News Recommendation with Deep Semantic Analysis

Cheng Chen, Xiangwu Meng, Zhenghua Xu, and Thomas Lukasiewicz

Abstract—With the popularity of mobile devices and the quick growth of the mobile Web, users can now browse news wherever they want; so, their news preferences are usually related to their geographical contexts. Consequently, many research efforts have been put on location-aware news recommendation, which recommends to users news happening nearest to them. Nevertheless, in a real-world context, users' news preferences are not only related to their locations, but also strongly related to their personal interests. Therefore, in this paper, we propose a hybrid method called location-aware personalized news recommendation with explicit semantic analysis (LP-ESA), which recommends news using both the users' personal interests and their geographical contexts. However, the Wikipedia-based topic space in LP-ESA suffers from the problems of high dimensionality, sparsity, and redundancy, which greatly degrade the performance of LP-ESA. To address these problems, we further propose a novel method called LP-DSA to exploit recommendation-oriented deep neural networks to extract dense, abstract, low dimensional, and effective feature representations for users, news, and locations. Experimental results show that LP-ESA and LP-DSA both significantly outperform the state-of-the-art baselines. In addition, LP-DSA offers more effective (19.8% to 179.6% better) online news recommendation with much lower time cost (25 times quicker) than LP-ESA.

Index Terms—Location-aware news recommendation, personalization, deep neural networks, deep semantic analysis.

I. INTRODUCTION

NOWADAYS, news reading is an indispensable daily activity of many people. With the recent popularity of smart mobiles and the rapid development of the mobile Web, more and more people tend to read news online via their mobiles or other handheld devices, e.g., tablets. However, due to the huge volume of news articles generated everyday, readers cannot afford to go through all the news online. So, news recommendation systems, which aim to filter out irrelevant online information and recommend to users their preferred news, have been widely studied [1], [2], [3], [4].

In classic personalized news recommendation systems, a user's news preferences are usually learned using his/her news reading history or other online activity histories; therefore, the user's news preferences are (almost) static in these systems. However, in real-world contexts, users' news preferences usually evolve with the change of their locations. For example,

C. Chen and X. Meng are with the Beijing Key Laboratory of Intelligent Telecommunications Software and Multimedia, Beijing University of Posts and Telecommunications, Beijing 100876, China.

C. Chen and X. Meng are with the School of Computer Science, Beijing University of Posts and Telecommunications, Beijing 100876, China.

Z. Xu and T. Lukasiewicz are with the Department of Computer Science, University of Oxford, UK.

Corresponding author: X. Meng (mengxw@bupt.edu.cn)

Manuscript received December 16, 2016; revised December 28, 2016.

people may prefer economic or political news, when they are working in the office; but they may like to read entertainment or sports news, when they are at home.

As the users' news preferences are strongly correlated with their geographical contexts, location-aware news recommendation systems that recommend news based on the geographical contexts of users have recently attracted many research efforts. There are mainly two directions for the research of location-aware news recommendation: physical distance-based and geographical topic-based.

Specifically, physical distance-based news recommendation [5], [6], [7], [8] aims to offer users with news happening nearest to them; so, the relevance of a news article to a user is measured by the physical distance between them based on GPS coordinates. However, the descriptions of event locations in many news articles are very vague and general (mentioning only a city or suburb) in practice; so, obtaining accurate GPS information for this kind of news is very difficult and sometimes even impossible, which greatly limits the application of physical distance-based methods.

Given this status quo, geographical topic-based methods [9], [10], [11] are proposed to achieve a more generic location-aware news recommendation without the need of accurate GPS information for news. The state-of-the-art geographical topic-based location-aware news recommendation method is *Explicit Localized Semantic Analysis (ELSA)* [11]. The recommendation process of ELSA is briefly as follows: it first uses collections of documents with geo-tags (called *geo-tagged documents*) as the descriptions of corresponding locations; then, it projects both the geo-tagged documents and the news articles onto a topic space using Explicit Semantic Analysis (ESA) [12], where Wikipedia concepts are regarded as topics; consequently, by considering link information between concepts, both locations (e.g., country, city, or venue) and news are represented as topic vectors and the relevance between a user and the candidate news is estimated by the similarities between the topic vectors of the user's current location and candidate news; finally, the news with top-k highest relevance are recommended to the user.

However, ELSA and most of the location-aware news recommendation methods take into account only the user's geographical information for news recommendation; so, the recommended news for different users at the same location will remain the same. This is, however, unreasonable for many real-world situations; for example, a housewife may prefer to read entertainment news at home, while a husband may like sports news. Therefore, users' news preferences strongly depend not only on their geographical contexts (i.e., locations), but also

on their personal interests; so, both should be considered to achieve satisfactory personalized news recommendations.

Motivated by this, we thus propose a hybrid news recommendation method, called *Location-aware Personalized news recommendation with Explicit Semantic Analysis* (abbreviated *LP-ESA*). LP-ESA offers personalized news recommendation based not only on the user's current location, but also on his/her personal interests, where a collection of geo-tagged documents is used as description of a location, while a collection of user history data (e.g., query history, browsing history, or tweeting history) is used to model the user's personal interests [13], [14].

Precisely, similarly to ELSA, LP-ESA first projects all the textual items (i.e., geo-tagged documents, user history data, and news articles) onto a Wikipedia-concept-based topic space using ESA [12]. Consequently, the users' interests and news are both modeled as weighted topic vectors, called *general user profiles* and *general news profiles*, respectively, whereas *local topic distributions* for locations are obtained by considering both the topic vectors of these locations and link information between the corresponding topics (Wikipedia concepts). Then, for each given user at a location, based on the general user profile and local topic distribution, LP-ESA constructs a *localized user profile* representing the probabilities of this user's news preferences (in terms of topics) at this location. Similarly, a *localized news profile* is also constructed for each news at a given location indicating the distribution of topics in the news at this location. Finally, the relevance of a news article to a user at a given location is measured by the similarity of this user's localized user profile and the localized news profile, and the news with top-k highest relevance scores are recommended. As LP-ESA utilizes both the geographical and preference information of users, it can achieve a much better news recommendation performance than ELSA.

However, as this is an ESA-based method, each Wikipedia concept is regarded as a topic in LP-ESA, and since the volume of concepts is enormous (millions) on Wikipedia, the resulting Wikipedia-based topic space in LP-ESA is very high-dimensional. Consequently, the process of online news recommendation in LP-ESA is very time consuming, which is unacceptable for the need of real-time response in practice. In addition, the Wikipedia-concept-based topic space also suffers from the problems of sparsity and redundancy (different concepts may have similar meanings, e.g., "university" and "college"), which degrade the news recommendation effectiveness of LP-ESA to a great extent.

We propose to address these problems and further improve the performance of LP-ESA by exploiting deep neural networks to map the (Wikipedia-concept-based) topic space to an abstract, dense, and low dimensional feature space, where the localized similarities between the users and their target news are maximized, and those with irrelevant news are minimized. We call this novel feature model *Deep Semantic Analysis* (abbreviated *DSA*) and the corresponding news recommendation method *Location-aware Personalized news recommendation with Deep Semantic Analysis* (abbreviated *LP-DSA*).

LP-DSA has the following advantages: (i) LP-DSA overcomes the huge dimensionality, sparsity, and redundancy prob-lems in LP-ESA by using deep neural networks to extract and compress the incoming Wikipedia-based topic features layer by layer and finally map them to a much denser abstract feature space with much lower dimensionality. (ii) The deep neural networks in LP-DSA are trained with a recommendation-oriented learning objective to differentiate the user's local target news from the irrelevant ones; so, the resulting abstract features for user, location, and news profiles are very effective representations for location-aware personalized news recommendation. (iii) Since the dimensionality of the resulting abstract feature space is much lower than that of the original Wikipedia-based topic space, LP-DSA greatly reduces the computational cost for online news recommendation. Consequently, LP-DSA is superior to LP-ESA in location-aware personalized news recommendation in terms of both recommendation effectiveness and efficiency, as shown by experimental studies.

In summary, the contributions of this paper are briefly as follows:

- We propose a hybrid news recommendation method, called location-aware personalized news recommendation with explicit semantic analysis (LP-ESA). LP-ESA innovatively takes into account both the geographical semantics of user's current location and his/her personal interests to achieve a better location-aware personalized news recommendation.
- We further propose a novel news recommendation method, called LP-DSA, to solve the huge dimensionality, sparsity, and redundancy problems in LP-ESA by deep semantic analysis. LP-DSA uses recommendation-oriented deep neural networks to extract dense, abstract, low dimensional, and effective feature representations for locations, users, and news, so as to improve the news recommendation performance in terms of both effectiveness and efficiency.
- We conduct extensive experiments based on a public real-world dataset to evaluate the news recommendation performances of the proposed LP-ESA and LP-DSA. The results show that LP-ESA and LP-DSA both significantly outperform the state-of-the-art baselines, ELSA [11] and STPM [9], in terms of precision at k, recall at k, F1-score at k, and truncated mean average precision at k. In addition, with the help of deep neural networks, LP-DSA further improves the news recommendation performance of LP-ESA: it offers more effective (19.8% to 179.6% better) news recommendation with much lower online recommendation time cost (25 times quicker) than LP-ESA.

The rest of this paper is organized as follows. We first review related works and give some preliminaries in Sections II and III, respectively. We then briefly introduce the ELSA method in Section IV, which is the most closely related state-of-the-art geographical topic-based location-aware news recommendation method. In Section V, we present the proposed location-aware personalized news recommendation method (LP-ESA) in detail, while Section VI introduces LP-DSA to exploit recommendation-oriented deep neural networks to

overcome the problems in LP-ESA and achieve a more effective and efficient news recommendation. Experimental studies and results are presented and evaluated in Section VII. Finally, Section VIII summarizes the paper and gives some concluding remarks and an outlook.

II. RELATED WORK

In this section, we discuss related works on news recommendation using personalized, location-aware, and location-aware personalized approaches (Sections II-A, II-B, and II-C, respectively). In addition, we also summarize works on personalized recommendation using deep learning techniques (Section II-D).

A. Personalized News Recommendation

Personalized news recommendation aims to recommend to users the news that match their personal interests best [1], [2]. Users' interests in news are usually modeled by their explicit ratings or browsing history (e.g., visited pages, reading times, and downloads) [15]. Both heuristic [4] and model-based methods [3], [16] are proposed for personalized news recommendation: the former are mainly based on mathematical or statistical solutions (e.g., cosine similarity and Euclidean distance), while the latter make use of machine learning techniques or mathematical models (e.g., Bayesian networks and decision trees). Specifically, Abel et al. [4] proposed to combine news with information on social media (tweets) to construct three kinds of user profiles, and then compute the cosine similarity between user profiles and news articles for personalized news recommendation. Yeung et al. [3] used Bayesian networks to predict levels of interesting news categories for users and to provide real-time personalized news recommendations, while a weighted-graph-based model is utilized in [16] to optimize the recommendation process.

B. Location-based News Recommendation

However, in the era of mobile and wireless networks, users' news preferences are also influenced by their geographical contexts, i.e., people usually pay more attention to the news that happened nearby than those far from them. Therefore, more and more research efforts have been put into location-aware news recommendation, which mainly focuses on two research directions: physical distance-based and geographical topic-based approaches.

As for physical distance-based news recommendation, Geo-Feed [5] and GeoRank [6] recommend to users some news happening at the users' current locations or within a given range, where GeoRank uses only static point locations of both users and news, while GeoFeed allows news with spatial extent; Pedro et al. [8] utilized the Euclidean distance between user and news locations to measure the importance of news; LocaNews [7] provides three versions of news and offers to users the most suitable ones according to their different distances to the news locations; Wen et al. [17] proposed a news stream recommendation framework, called MobiFeed, to further study news recommendation according to users' moving tracks.

However, in the real-world context, the description of event locations in many news articles are very vague and general (mentioning only a city or suburb); so, obtaining accurate GPS information for this kind of news is very difficult and sometimes even impossible. Consequently, the application of physical distance-based methods is heavily limited.

As for more generic location-aware news recommendation, geographical topic-based methods are proposed. Instead of using GPS coordinates, the locations are described using topic vectors, and the relevance of a news article to a user is measured by the similarity between the topic vectors of the news and the current location of the user. Therefore, the topic representations of locations are crucial for topic-based location-aware news recommendation, and a range of topic models (such as Latent Dirichlet Allocation (LDA) [18], [19], Explicit Semantic Analysis (ESA) [12], Probabilistic Latent Semantic Analysis (PLSA) [20], and their improved models [9], [10], [11], [21]) have been used for geographical topic-based location-aware news recommendation.

The state-of-the-art geographical topic-based location-aware news recommendation method is *Explicit Localized Semantic Analysis (ELSA)* [11], which is reported to outperform many other geographical topic-based methods. ELSA is an ESA-based method and is the most similar location-aware news recommendation solution to this work. But ELSA is different from our work in the following aspects: (i) The proposed LP-ESA and LP-DSA in this paper both capture not only the location information but also the personal interests of users for news recommendation; while ELSA is solely based on the location. (ii) Neural networks are used to solve the high dimensionality, sparsity, and redundancy problems in LP-ESA; as an ESA-based method, these problems also exist in ELSA, but they are not properly addressed. Thus, our proposed LP-ESA and LP-DSA methods achieve much better performance than ELSA.

C. Location-aware Personalized News Recommendation

Most of the existing location-aware news recommendation methods are based on location information only; so, they recommend news solely based on users' locations, but regardless of the users' personal interests (i.e., non-personalized). Thus, different users in the same location will receive the same recommendation, which is unreasonable in practise.

To our knowledge, the Spatial Topical Preference Model (STPM) [9] is the only location-aware news recommendation system that considers both the locations and personal interests of users. However, there also exist some differences between STPM and our LP-ESA. (i) STPM is an LDA-based method; so, all textual content in STPM is represented by latent topics, while LP-ESA is ESA-based; so, texts are represented by a Wikipedia-based explicit topic space. (ii) More importantly, STPM has a cold start problem: the localized user profiles in STPM are directly learned using the users' history data on the corresponding locations (e.g., home or office); so, it fails when users travel to a new place; to avoid this problem, the localized user profiles in LP-ESA are inferred from the users' general profiles (modeled with users' entire history data) and

the local topic distributions of locations; therefore, LP-ESA overcomes the cold start problem and is also applicable to recommendation at new locations.

D. Personalized Recommendation Using Deep Learning

Due to its capability to extract effective representations [22], deep learning has already been successfully applied in many personalized recommendation applications, such as music recommendation [23], movie recommendation [24], tag-aware recommendation [25], and multi-view item recommendation [26].

Similarly to our work, [26] also uses deep neural networks with a ranking-oriented training objective for dimensionality reduction. However, [26] is very different from the LP-DSA proposed here: (i) [26] is not a location-aware model; so, its recommendation will not be sensitive to the changes of users' geographical contexts; (ii) [26] is not an ESA-based method; so, the ranking-oriented neural networks are not used to solve the huge dimensionality, sparsity, and redundancy problems in the Wikipedia-concept-based topic space. (iii) [26] has to train own parameters for each neural network, while in our work, shared parameters are applied in deep neural networks, because all input topic vectors in our work share the same Wikipedia-based topic space; consequently, the time needed for model training in our work is greatly reduced.

III. PRELIMINARIES

To achieve personalized recommendation, users' personal interests are usually learned from their history data on the Web (e.g., query history, browsing history, or tweeting history). In this paper, the tweeting histories of users are utilized to learn their personal interests on news. This is because, (i) as studied in [4], most tweets (over 85%) posted everyday are related to the latest news; (ii) tweets usually represent the personal opinions of users; so, their interests can be harvested from their tweeting histories; and (iii) tweets are posted publicly by users; so, they are usually freely accessible and contain little sensitive information about their creators, such that using tweeting histories is more privacy-enhanced than using other history data (e.g., browsing or query history).

Consequently, given U, V, L, and Z as the sets of *users*, *news articles*, *locations*, and *topics*, respectively, we have the following definitions.

If a user u posts, retweets, or comments on a tweet containing an URL to a news article v^* at a location l, then v^* is a *target news* of u at l for location-aware personalized news recommendation.

According to the explicit semantic analysis model [12], each *topic* $z \in Z$ is a Wikipedia concept, which can be semantically represented as a word distribution θ_z on the related Wikipedia article.

For each location $l \in L$, a *local topic distribution* θ_l is used by ELSA and LP-ESA to capture the importance of topics Z at l: θ_l is modeled using a set D_l of documents with geo-tags (called *geo-tagged documents*), where geo-tags are location names or GPS coordinates; and the Google Maps API[1] can be

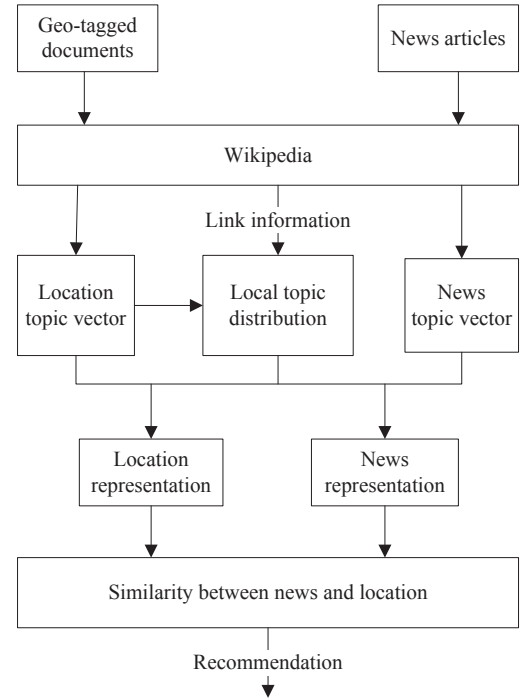[1] https://developers.google.com/maps/

Fig. 1. The overall process of ELSA

used to translate the coordinates to the corresponding location names.

Both *user* and *news profiles* in LP-ESA are topic vectors, represented as the probability distributions over topics. Specifically, for each user $u \in U$, a *general user profile* φ_u is used to capture u's general news preferences, which is constructed from u's (tweeting) history data H_u. Since the general user profile is static for u at different locations, to capture u's localized news preferences at a given location $l \in L$, a *localized user profile* $\Phi_{u,l}$ is constructed based on u's general user profile φ_u and l's local topic distribution θ_l. Similarly, for each news v, a *general news profile* φ_v is constructed based on the corresponding news article, while a *localized news profile* $\Phi_{v,l}$ is obtained using φ_v and θ_l.

Given a set of news V and a user u at location l, *location-aware personalized news recommendation* is to generate a ranked recommendation list $\{v_1 \succeq v_2 \succeq \cdots \succeq v_{|V|}\}$ for all news, such that $v_i \succeq v_j$ iff $R_{u,v_i,l} \geq R_{u,v_j,l}$, where $R_{u,v,l}$ is a score measuring the relevance of news v to user u at location l, which is computed by the cosine similarity of the localized user profile $\Phi_{u,l}$ and the localized news profile $\Phi_{v,l}$.

IV. EXPLICIT LOCALIZED SEMANTIC ANALYSIS IN NEWS RECOMMENDATION

Explicit localized semantic analysis (ELSA) [11] is the state-of-the-art topic model used for geographical topic-based location-aware news recommendation, which is reported to outperform many other topic models, e.g., BOW, LDA, and ESA. Due to its close relation to our work, we first briefly present ELSA in this section.

ELSA is an ESA-based [12] solution, where each Wikipedia concept is considered as a potential topic and each location and news is represented as a Wikipedia-based topic vector.
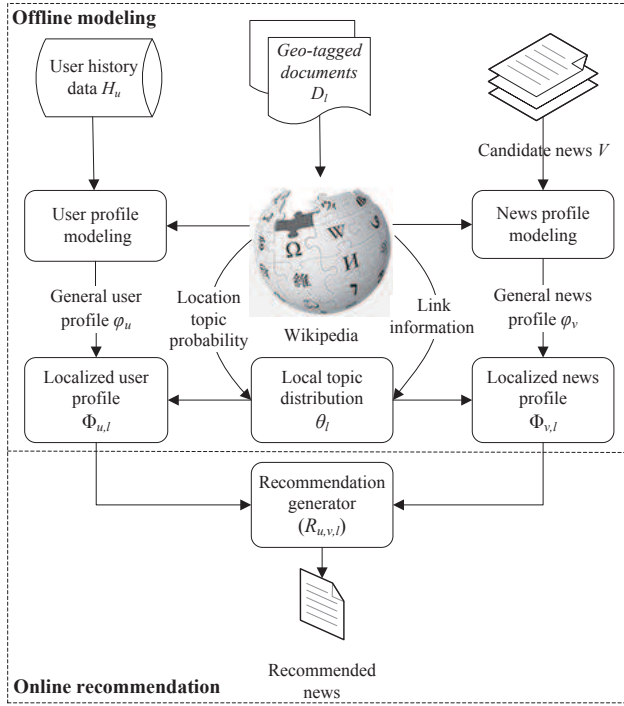
Fig. 2. The overall process of LP-ESA

Figure 1 shows the overall process of ELSA. First of all, ELSA collects for each location a set of documents with the corresponding geo-tags as the description of this location. Then, these geo-tagged documents and the candidate news articles are mapped onto a Wikipedia-based topic space to generate for each location or news a topic vector, which is represented as a probability distribution over topics and called *location topic vector* or *news topic vector*, respectively. Consequently, the topics related to a location are the ones with non-zero probability values in this location's topic vector. Since these topics generally depend on each other, ELSA further uses the link information within the corresponding Wikipedia concepts to construct a topic dependency graph and then applies PageRank [27] to estimate a *local topic distribution*. With the help of the local topic distribution, the location and the news topic vectors are localized to obtain the local topic representations of locations and news, which are used to estimate the similarities between news and locations. Finally, the recommendation is made by offering to the user the news with top-k similarity scores to his/her current location.

V. LOCATION-AWARE PERSONALIZED NEWS RECOMMENDATION WITH EXPLICIT SEMANTIC ANALYSIS

Since ELSA and most of the location-aware news recommendation methods consider solely the user's location for news recommendation; the recommended news for different users at the same location will remain the same. However, in practice, users' news preferences are strongly correlated not only to their geographical contexts (i.e., locations) but also to their personal interests, so both should be considered to achieve satisfactory personalized news recommendations.

Therefore, we propose a hybrid news recommendation method, called location-aware personalized news recommen-
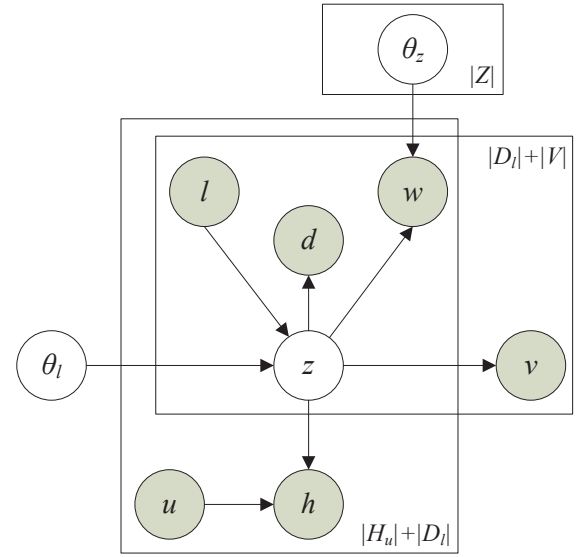

Fig. 3. Graphical representation of LP-ESA

dation with explicit semantic analysis (LP-ESA), which takes into account both the user's current location and also his/her personal interests. Consequently, differently from ELSA, in LP-ESA, the user's localized profile is changing over locations such that different users at a same location will have different and personalized news recommendations.

Generally, LP-ESA is a personalized extension of ELSA and the overall process of LP-ESA is depicted in Figure 2, which generally consists of two steps: offline modeling and online recommendation. Offline modeling aims to construct the localized profiles for users and news, which are then used to conduct location-aware personalized news recommendation in the online recommendation step.

Specifically, in the offline modeling step, given a user u, LP-ESA first projects the user's history data H_u and each news article v in the news set V onto a Wikipedia topic space to construct a *general user profile* φ_u for the given user u and a *general news profile* φ_v for each news v, respectively. However, both profiles are insensitive to the change of the user's geographical context. Therefore, to achieve location-aware personalized news recommendation, LP-ESA collects a set of geo-tagged documents D_l for a given location l, and estimates a local topic distribution θ_l using l's geographical topics and their corresponding link information as done in ELSA; then, LP-ESA models a *localized user (resp., news) profile* $\Phi_{u,l}$ (resp., $\Phi_{v,l}$) for each user u (resp., news v) at location l, according to both the general user (resp., news) profile φ_u (resp., φ_v) and the local topic distribution of l (i.e., θ_l). Finally, both localized user and news profiles are used in the online recommendation phase to generate for the given user u a ranked list of news according to the relevancy score $R_{u,v,l}$, computed by the cosine similarity of $\Phi_{u,l}$ and $\Phi_{v,l}$. Formally,

$$R_{u,v,l} = Sim(\Phi_{u,l}, \Phi_{v,l}) = \frac{\Phi_{u,l} \cdot \Phi_{v,l}}{\|\Phi_{u,l}\| \cdot \|\Phi_{v,l}\|}. \quad (1)$$

The graphical representation of LP-ESA is shown in Figure 3, where a geo-tagged document $d \in D_l$, a history data

record $h \in H_u$, a news article $v \in V$, and a word w are all dependent on topic $z \in Z$, while the topics depend on a location l. In addition, l, u, d, h, and v are observed variables, while z, θ_z, and θ_l are the unobserved ones.

A. General User Profile

As the key of personalized news recommendation, for a given user, a user profile is used to describe this user's personal interests in news, which can be modeled using his/her history data on the Web (e.g., query history, browsing history, or tweeting history). Therefore, LP-ESA first collects a set of tweeting history data H_u for each given user u and then projects H_u onto the Wikipedia topic space to model u's general user profile (denoted φ_u); formally,

$$\varphi_u = \langle p(H_u|z_1, \theta_{z_1}), \ldots, p(H_u|z_i, \theta_{z_i}), \ldots, p(H_u|z_{|Z|}, \theta_{z_{|Z|}}) \rangle,$$

where $p(H_u|z_i, \theta_{z_i})$ is the probability of the topic z_i in the tweeting history data H_u of u. Then, given $h \in H_u$ as a tweeting record, we have

$$p(H_u|z_i, \theta_{z_i}) = \prod_{h \in H_u} p(h|z_i, \theta_{z_i})$$
$$\propto \sum_{h \in H_u} \log p(h|z_i, \theta_{z_i}).$$

where $p(h|z_i, \theta_{z_i})$ is the probability of a topic z_i in the tweeting record h.

Differently from ELSA, where the probability of a topic in a document (e.g., news article) is estimated solely based on the possibility that the words in the document are generated from this topic, LP-ESA considers both the probability of generating a tweeting record h from each word in h (i.e., $p(h|w)$) and that of the word w generated from each topic z_i (i.e., $p(w|z_i, \theta_{z_i})$) to estimate the probability of a topic z_i in the tweeting record h (i.e., $p(h|z_i, \theta_{z_i})$). Consequently, the word with higher probability in h contributes more to the topic probability estimation in LP-ESA, while all words of the document are equally important in ELSA. Therefore, assuming all words are independent, we have the following derivation:

$$\log p(h|z_i, \theta_{z_i}) = \sum_{w \in h} \log p(h|w) p(w|z_i, \theta_{z_i}), \quad (2)$$

where both $p(h|w)$ and $p(w|z_i, \theta_{z_i})$ are estimated using a variant of TF-IDF [28] in LP-ESA; formally,

$$p(h|w) \approx \frac{(1 + \log n_h(w)) * \log \frac{|Z|}{n_Z(w)}}{\sqrt{\sum_{w_j \in h} [(1 + \log n_h(w_j)) * \log \frac{|Z|}{n_Z(w_j)}]^2}}, \quad (3)$$

$$p(w|z_i, \theta_{z_i}) \approx \frac{(1 + \log(n_{r_i}(w) + 1)) * \log \frac{|Z|}{n_Z(w)}}{\sqrt{\sum_{w_j \in r_i} [(1 + \log(n_{r_i}(w_j) + 1)) * \log \frac{|Z|}{n_Z(w_j)}]^2 + \log |Z|}}, \quad (4)$$

where $n_h(w)$ and $n_{r_i}(w)$ denote the frequency of w appearing in h and r_i (a Wikipedia article with regard to topic z_i), respectively; and $n_Z(w)$ is the number of topics in Z whose related Wikipedia articles contain w.

B. General News Profile

Similarly, LP-ESA also models a general news profile (φ_v) for each news article v; formally,

$$\varphi_v = \langle p(v|z_1, \theta_{z_1}), \ldots, p(v|z_i, \theta_{z_i}), \ldots, p(v|z_{|Z|}, \theta_{z_{|Z|}}) \rangle,$$

where $p(v|z_i, \theta_{z_i})$ is the probability of topic z_i in the news article v. Specifically, LP-ESA utilizes both the probability of obtaining a news article v given each word w in v (i.e., $p(v|w)$) and that of each word w in v generated from each topic z_i (i.e., $p(w|z_i, \theta_{z_i})$) to estimate $p(v|z_i, \theta_{z_i})$, while ELSA merely considers the latter factor. Therefore, Similar to Formula (2), $p(v|z_i, \theta_{z_i})$ is estimated as follows:

$$p(v|z_i, \theta_{z_i}) = \prod_{w \in v} p(v|w) p(w|z_i, \theta_{z_i}) \propto \sum_{w \in v} \log p(v|w) p(w|z_i, \theta_{z_i}),$$

where $p(w|z_i, \theta_{z_i})$ can be estimated by Formula (4), and $p(v|w)$ can be estimated based on a variant of TF-IDF as follows,

$$p(v|w) \approx \frac{(1 + \log n_v(w)) * \log \frac{|Z|}{n_Z(w)}}{\sqrt{\sum_{w_j \in v} [(1 + \log n_v(w_j)) * \log \frac{|Z|}{n_Z(w_j)}]^2}}.$$

C. Local Topic Distribution

To achieve location-aware personalized news recommendation, LP-ESA estimates a local topic distribution for each location l as done in ELSA [11]. Formally, a local topic distribution of location l is defined as

$$\theta_l = \langle p(z_1|l, \theta_l), \ldots, p(z_i|l, \theta_l), \ldots, p(z_{|Z_l|}|l, \theta_l) \rangle,$$

where z_i is a topic related to location l, and $p(z_i|l, \theta_l)$ is the probability of topic z_i at l.

To identify all the topics related to l, LP-ESA first collects a set of geo-tagged documents (i.e., D_l) for location l; then, LP-ESA projects D_l onto the Wikipedia topic space to learn the topics and their probabilities, where the topics with non-zero probabilities at l are believed to be the ones actually related to l (i.e., local topics). However, most topics at a location l are coherent and depend on each other. Therefore, LP-ESA constructs a dependency structure over local topics of l to simplify the estimation of $p(z_i|l, \theta_l)$, where the dependency is based on the links between the corresponding concepts of the local topics in Wikipedia. Finally, $p(z_i|l, \theta_l)$ is approximated by applying the PageRank algorithm [27] to this structure. Formally,

$$p(z_i|l, \theta_l) \approx PageRank(z_i)$$
$$= \frac{1-q}{|Z_l|} + \sum_{z_j \in in(z_i, Z_l)} \frac{PageRank(z_j)}{L_{Z_l}(z_j)}, \quad (5)$$

where q is a damping factor, $in(z_i, Z_l)$ is the set of topics in Z_l that has links to z_i, and $L_{Z_l}(z_j)$ is the out-degree of z_j.
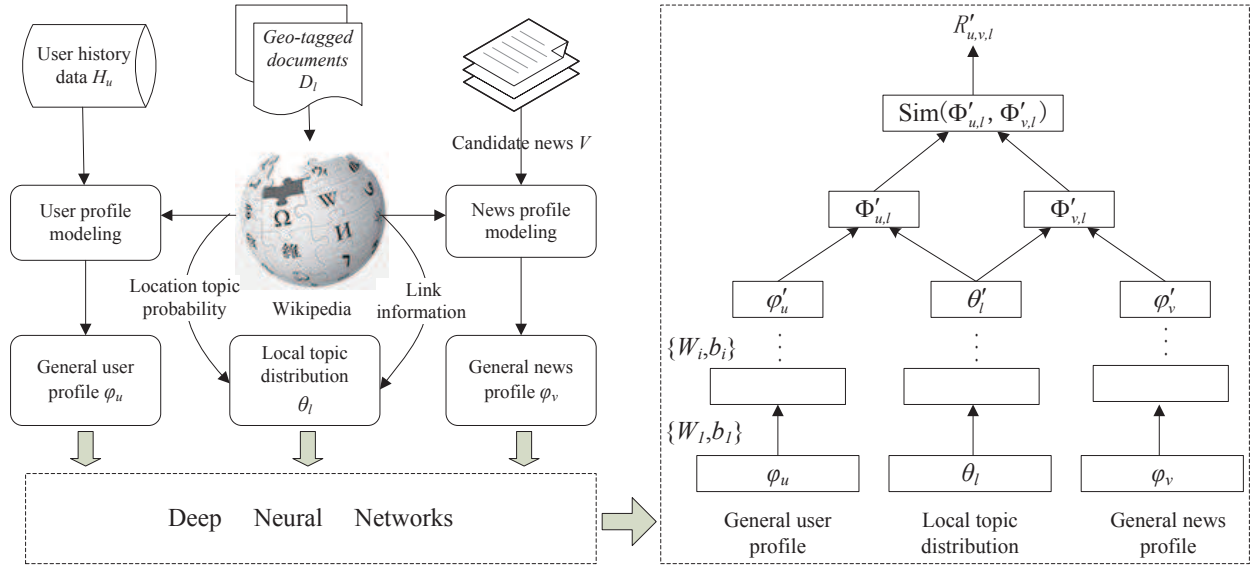
Fig. 4. The overall process of location-aware personalized news recommendation with deep semantic analysis

D. Localized User Profile

As the users' news preferences usually evolve with the change of their locations, for a given user u at a location l, LP-ESA further constructs a dynamic localized user profile based on u's general user profile φ_u and the local topic distribution θ_l of l. Formally, the localized user profile of user u at location l is defined as

$$\Phi_{u,l} = \langle p(H_u, z_1, l|\theta_l, \theta_{z_1}), \ldots, p(H_u, z_i, l|\theta_l, \theta_{z_i}), \ldots, p(H_u, z_{|Z|}, l|\theta_l, \theta_{z_{|Z|}}) \rangle,$$

where $p(H_u, z_i, l|\theta_l, \theta_{z_i})$ measures the probability of u's preferences relative to topic z_i at location l. Since z depends on l and θ_l, while l is independent of θ_l and θ_z (as shown in Figure 3), by assuming $p(l)$ to be uniform, we then estimate $p(H_u, z_i, l|\theta_l, \theta_{z_i})$ as follows:

$$\begin{aligned} p(H_u, z_i, l|\theta_l, \theta_{z_i}) &= p(H_u|z_i, \theta_{z_i}) \cdot p(z_i|l, \theta_l) \cdot p(l) \\ &\propto p(H_u|z_i, \theta_{z_i}) \cdot p(z_i|l, \theta_l) \qquad (6) \\ &\propto \sum_{h \in H_u} \log p(h|z_i, \theta_{z_i}) + \log p(z_i|l, \theta_l), \end{aligned}$$

where $p(z_i|l, \theta_l)$ is the local topic probability in the local topic distribution θ_l, and $p(H_u|z_i, \theta_{z_i})$ is the topic probability in the general user profile φ_u.

E. Localized News Profile

Based on the general news profile φ_v and the local topic distribution θ_l at a location l, the localized news profile of a news article v at l is defined as follows:

$$\Phi_{v,l} = \langle p(v, z_1, l|\theta_l, \theta_{z_1}), \ldots, p(v, z_i, l|\theta_l, \theta_{z_i}), \ldots, p(v, z_{|Z|}, l|\theta_l, \theta_{z_{|Z|}}) \rangle,$$

where $p(v, z_i, l|\theta_l, \theta_{z_i})$ is the probability of topic z_i in news v at l. Similarly to Formula (6), we have:

$$\begin{aligned} p(v, z_i, l|\theta_l, \theta_{z_i}) &= p(v|z_i, \theta_{z_i}) \cdot p(z_i|l, \theta_l) \cdot p(l) \\ &\propto p(v|z_i, \theta_{z_i}) \cdot p(z_i|l, \theta_l), \qquad (7) \end{aligned}$$

where $p(v|z_i, \theta_{z_i})$ is the topic probability in the general news profile φ_v, and $p(z_i|l, \theta_l)$ is the local topic probability in θ_l.

VI. LOCATION-AWARE PERSONALIZED NEWS RECOMMENDATION WITH DEEP SEMANTIC ANALYSIS

By taking into account both the geographic information and personal interests of users, LP-ESA greatly outperforms ELSA in news recommendation (as demonstrated in the experimental studies in Section VII). However, as an ESA-based method, each Wikipedia concept in LP-ESA is regarded as a topic, and since there is a high volume of concepts on Wikipedia, the dimensionality of the resulting Wikipedia-based topic space in LP-ESA is very large. Consequently, the online recommendation phase in LP-ESA that has to compute the cosine similarities between the high dimensional localized profiles of the given user and all candidate news is very time-consuming, which is unacceptable for the need of real-time response in practice. In addition, the Wikipedia-concept-based topic space also suffers from the problems of sparsity and redundancy (different concepts may have similar meanings, e.g., "university" and "college"), which degrade the news recommendation effectiveness of LP-ESA to a great extent.

To solve these problems and further improve the performance of LP-ESA, we thus propose a novel news recommendation method, called location-aware personalized news recommendation with deep semantic analysis (LP-DSA), which utilizes deep neural networks to map the Wikipedia-based topic space to an abstract, dense, and low dimensional feature space, where the localized similarities between the users and their target news are maximized, and those with irrelevant news are minimized.

Figure 4 depicts the overall process of LP-DSA. First of all, LP-DSA takes the general user profile φ_u, local topic distribution θ_l, and general news profile φ_v (generated in LP-ESA) as the respective inputs of three deep neural networks with shared parameters. Then, these inputs are then passed through multiple hidden layers and projected onto an abstract, dense, and low dimensional feature space on the final hidden layer, resulting in three abstract general feature representations for the user, location, and news, which are called *abstract*

general user profile (denoted φ'_u), *abstract local topic distribution* (denoted θ'_l), and *abstract general news profiles* (denoted φ'_v), respectively. Then, LP-DSA models an *abstract localized user profile* (denoted $\Phi'_{u,l}$) using φ'_u and θ'_l and an *abstract localized news profile* (denoted $\Phi'_{v,l}$) using φ'_v and θ'_l. Finally, for a given user, LP-DSA recommends to this user the news with top-k highest relevance scores, which are computed by applying the softmax function on the cosine similarities of the corresponding abstract localized profiles of the user and the news. To learn the parameters W_i and b_i in these deep neural networks, the model is trained using an objective to maximize the softmax similarities between the users and their target news, while minimizing those with irrelevant ones.

Formally, given a weight matrix W_1 and a bias vector b_1, LP-DSA takes the general user profile φ_u, local topic distribution θ_l, and general news profile φ_v as the inputs and generates the intermediate outputs f_1 of the first hidden layer according to the following formulas:

$$f_1(u) = \tanh(W_1\varphi_u + b_1), \tag{8}$$
$$f_1(l) = \tanh(W_1\theta_l + b_1), \tag{9}$$
$$f_1(v) = \tanh(W_1\varphi_v + b_1), \tag{10}$$

where $\tanh$ is adopted as the activation function. Similarly, the intermediate outputs of the i-th hidden layer ($i = 2, \ldots, N$) are computed by

$$f_i(u) = \tanh(W_i f_{i-1}(u) + b_i), \tag{11}$$
$$f_i(l) = \tanh(W_i f_{i-1}(l) + b_i), \tag{12}$$
$$f_i(v) = \tanh(W_i f_{i-1}(v) + b_i), \tag{13}$$

where W_i and b_i are the weight matrix and the bias vector for the i-th hidden layer, and N is the total number of hidden layers. Consequently, the outputs of the N-th hidden layer are the abstract general user profile φ'_u, the abstract local topic distribution θ'_l, and the abstract general news profile φ'_v. Formally,

$$\varphi'_u = f_N(u), \quad \theta'_l = f_N(l), \quad \varphi'_v = f_N(v). \tag{14}$$

Furthermore, since φ'_u, θ'_l, and φ'_v share the same feature space, according to Formulas (6) and (7), the abstract localized user and news profiles can be formally defined as

$$\Phi'_{u,l} = \varphi'^T_u \cdot \theta'_l, \tag{15}$$
$$\Phi'_{v,l} = \varphi'^T_v \cdot \theta'_l. \tag{16}$$

Then, the similarity between a user and a news article at location l is measured using the cosine similarity between their abstract localized profiles at l; formally,

$$Sim(\Phi'_{u,l}, \Phi'_{v,l}) = \frac{\Phi'_{u,l} \cdot \Phi'_{v,l}}{\|\Phi'_{u,l}\| \cdot \|\Phi'_{v,l}\|}. \tag{17}$$

Finally, LP-DSA recommends news articles based on their relevance scores to a given user u at location l, where the relevance scores are computed by applying the softmax function on the resulted similarities; formally,

$$R'_{u,v,l} = \frac{e^{Sim(\Phi'_{u,l}, \Phi'_{v,l})}}{\sum_{v' \in V} e^{Sim(\Phi'_{u,l}, \Phi'_{v',l})}}, \tag{18}$$

Therefore, to achieve a good personalized news recommendation, the target news should have higher relevance scores than others. We thus conduct the model training with an objective to maximize the relevance scores of target news; equivalently, this means to maximize the localized similarities between users and their target news, and to minimize those with irrelevant news. Formally, this is equivalent to minimize the following loss function:

$$
\begin{aligned}
Loss(\Theta) &= -\sum_{(u,v^*,l)} \log(R'_{u,v^*,l}) \\
&= -\sum_{(u,v^*,l)} [\log(e^{Sim(\Phi'_{u,l}, \Phi'_{v^*,l})}) \\
&\quad - \log(\sum_{v' \in V} e^{Sim(\Phi'_{u,l}, \Phi'_{v',l})})],
\end{aligned} \tag{19}
$$

where Θ represents the set of parameters $\{W_i, b_i\}$ ($i = 1, \ldots, N$) in the neural networks; the tuple (u, v^*, l) is a training sample, which represents that a user u has a target news article v^* at the location l, which is generated from the user's tweeting history records in the training dataset.

According to the above loss function, ideally, all news in V should be used as candidate news in the model training. However, due to the large number of candidate news, this will result in a very expensive training cost. Therefore, in practice, negative sampling is adopted to train the model more efficiently and to ensure the scalability of LP-DSA [29]. For each training sample (u, v^*, l), we approximate V by including the target news v^* and only K irrelevant news v^- randomly sampled from candidate set V, where (u, v^-, l) is called negative sample. Therefore, the loss function with negative sampling is formally defined as

$$
\begin{aligned}
Loss^{NS}(\Theta) &= -\sum_{(u,v^*,l)} [\log(e^{Sim(\Phi'_{u,l}, \Phi'_{v^*,l})}) \\
&\quad - \log(\sum_{v' \in V^-} e^{Sim(\Phi'_{u,l}, \Phi'_{v',l})})],
\end{aligned} \tag{20}
$$

where V^- is an approximative candidate set. Our pilot study has shown that negative sampling significantly reduces the training cost, and only marginally damages the training effectiveness [29].

The training process is conducted as follows: we first initialize the weight matrices W_i using the random normal distribution, and initialize the biases b_i to 0; the model is then trained via stochastic gradient descent [30], which is a gradient-based optimization algorithm; finally, the training stops when the model converges or reaches the maximum training iterations.

VII. EXPERIMENTAL STUDIES

A. Baselines

To show the strength of our proposed methods, LP-ESA and LP-DSA, in geographical topic-based location-aware news recommendation, the following two state-of-the-art methods based on geographical topics are selected as the baselines:

- **Explicit Localized Semantic Analysis (ELSA)** [11]: as presented in Section IV, ELSA is the state-of-the-art topic-based location-aware news recommendation

TABLE I
STATISTIC INFORMATION OF DATASET

Tweets	Users	News	Locations	Samples
2,316,204	1,619	63,485	2,366	98,321

TABLE II
DETAILS OF TRAINING SET AND TEST (SUB)SETS

	Users	News	Locations	Samples
Training set	1,555	53,036	2,295	78,946
Test set	909	15,853	743	19,375
Old City test subset	825	12,518	325	14,883
New City test subset	788	3,822	662	4,492

method, which is reported to outperform many other geographical topic-based methods. Since LP-ESA is a personalized extension of ELSA, by comparing to ELSA, the strength of taking into account the personal interests of users in location-aware news recommendation can be demonstrated.

- **Spatial Topical Preference Model (STPM)** [9]: as introduced in Section II-C, to our knowledge, STPM is the only geographical topic-based news recommendation system that considers both the locations and personal interests of users. However, instead of using Explicit Semantic Analysis (ESA), the topic model used in STPM is Latent Dirichlet Allocation (LDA) [18]. As the state-of-the-art and only solution for the location-aware personalized news recommendation, STPM is also selected as baseline.

B. Dataset and Preprocessing

Our experimental study is based on a publicly available Twitter dataset [4], which consists of $2,316,204$ tweets posted by $1,619$ users ($|U| = 1,619$). About half (over one million) of these tweets explicitly contain URLs to the news articles; by using these URLs to download the corresponding news articles, we get $63,485$ news articles, which are used as the candidate news articles for recommendation, i.e., $|V| = 63,485$. Then, we apply a Web service tool[2] to extract city names from the news articles, resulting in $2,366$ locations ($|L| = 2,366$). Finally, we consider these city names as geo-tags and use the titles and keywords of the news articles, from which the city names are extracted, as the descriptions of these locations, i.e., geo-tagged documents D_l. The statistic information of the dataset is summarized in Table I.

To model the users' news preferences, if a user u posts a tweet which has an URL to a news article v^* containing a city name (location) l, u is believed to be interested in v^* at l, from which a sample (u, v^*, l) is generated, indicating v^* is a target news of u at l. Consequently, a total of $98,321$ samples are obtained from the dataset; we randomly select 80% of the samples as the training set and the remaining 20% as the test set. We further divide the test set into two subsets:

[2]OpenCalais at https://opencalais.com/.

Old City test subset and New City test subset, to evaluate the methods' news recommendation performance under different geographical contexts. We do so, as it is generally more difficult to achieve satisfactory news recommendations in practice, when users are at new locations, which they have not visited before and for which there are no local history data. Specifically, for each sample (u, v^*, l) in the test set, if the training set also contains some samples related to the same user u at the same location l, l is seen as a city (location) that u has visited before; so, (u, v^*, l) is added to the Old City test subset; otherwise, l is a new city to user u, and the sample (u, v^*, l) is added to the New City test subset. The details of the training set and test (sub)sets are summarized in Table II.

Furthermore, a news recommendation system always aims to recommend to users news that they have never read before. Therefore, to be in line with this requirement, we remove the tweets where test samples are extracted and use only the rest of the tweeting records in the dataset as the users' history data (i.e., H_u) to construct user profiles. Finally, a Wikipedia snapshot of August 11, 2014 is used for the explicit semantic analysis [12] in ELSA, LP-ESA, and LP-DSA, resulting in $1,301,900$ concepts with $1,618,970$ distinct terms. To cut down the calculation and the memory cost, we select $8,000$ most frequent topics/concepts among all the resulting topic representations of the users, locations, and news as the shared Wikipedia topic space.

C. Implementation Settings and Evaluation Metrics

All methods are implemented using Python and Theano (a Python library for machine learning) and run on a GPU server of Oxford University's ARC facility [31] with an NVIDIA Tesla K20 and 8GB graphics card memory. The related parameters of STPM are set according to [9]; the damping factor q in ELSA, LP-ESA, and LP-DSA is set to 0.85. Other parameters in LP-DSA are set empirically as follows:

- Number of hidden layers: $N = 3$.
- Number of neurons in the first, second, and third hidden layer: 1024, 512, and 256, respectively;
- Number of news articles that are randomly selected to approximate V for each training sample: $K = 127$;
- Learning rate in model training: 0.0001.

As for the evaluation of recommendation systems, the most popular metrics are precision, recall, and F1-score [32]. Since most users usually only browse the topmost recommended news, we apply these metrics at a given cut-off rank k, i.e., considering only the top-k results on the recommendation list, called *precision at k (P@k)*, *recall at k (R@k)*, and *F1-score at k (F@k)*. Formally,

$$P@k = \frac{1}{|U'|} \sum_{u \in U'} P_u@k, \qquad P_u@k = \frac{C_u@k}{k}, \qquad (21)$$

$$R@k = \frac{1}{|U'|} \sum_{u \in U'} R_u@k, \qquad R_u@k = \frac{C_u@k}{N_u}, \qquad (22)$$

$$F@k = \frac{2 * P@k * R@k}{P@k + R@k}, \qquad (23)$$

where k is the length of the recommendation list; U' is the set of users in the test (sub)sets; $P_u@k$ and $R_u@k$ are the precision and recall at k for a given user u, respectively; $C_u@k$

is the number of u's target news in the recommendation list; and N_u is the total number of u's target news in test (sub)sets.

Although precision, recall, and F1-score are very useful metrics, they do not consider the positions of the target news in a recommendation list. In practice, users always prefer to have their target items ranked as top as possible in a recommendation list, especially for the news reading on the mobile devices whose display space is very limited. Therefore, we also employ the truncated mean average precision at k ($MAP@k$) [33] as evaluation metric to take into account the order of news and give higher weights to those ranked in front. Formally,

$$AP_u@k = \frac{\sum_{i=1}^{k}(P_u@i * rel@i)}{\min(k, N_u)},$$
$$MAP@k = \frac{1}{|U'|}\sum_{u \in U'} AP_u@k, \qquad (24)$$

where $AP_u@k$ is the average precision for a given user u, and $rel@i$ is an indicator function, which is 1, if the i-th news on the recommendation list is a target news for u, and 0, otherwise.

D. Recommendation Performance

Main results. Figure 5 depicts the news recommendation effectiveness of LP-DSA, LP-ESA, ELSA, and STPM on three test (sub)sets (i.e., Old City test subset, New City test subset, and the whole test set) in terms of precision at k ($P@k$), recall at k ($R@k$), F1-score at k ($F@k$), and truncated mean average precision at k ($MAP@k$), where the value of k varies from 1 to 20. As shown in Figure 5, LP-ESA and LP-DSA both significantly outperform the state-of-the-art baselines, ELSA and STPM, in terms of all metrics; moreover, with the help of deep semantic analysis, LP-DSA greatly improves the recommendation performance of LP-ESA.

Figures 5 (a)-(c) exhibit the precision at k ($P@k$) of all the four methods on three test (sub)sets. Specifically, Figure 5 (a) presents the results on the Old City test subset; it is shown that LP-DSA achieves the best performance in $P@k$ for all k values, while LP-ESA also outperforms ELSA and STPM constantly. The values of $P@k$ for all methods are relatively stable with the change of k with a peak value for LP-DSA when $k = 2$. Figure 5 (b) depicts the $P@k$ results of LP-DSA, LP-ESA, and ELSA on the New City test subset, where the general tendency and the comparison results between these three methods are similar to those in Figure 5 (a). However, STPM is inapplicable to the New City test subset, because of the cold start problem: given a user at a location, STPM requests the user's history data on this specific location to model localized news preferences; however, for each sample in the New City test subset, there is no sample in the training set sharing both the same location and same user with it; so STPM is unable to construct the localized news preferences for users in the New City test subset and, consequently, unable to conduct news recommendation on the New City test subset. Figure 5 (c) witnesses the results on the whole test set, where similar conclusions to Figure 5 (a) can be drawn.

Figures 5 (d)-(f) show the methods' recall at k ($R@k$) on Old City test subset, the New City test subset, and the whole

test set, respectively. Similarly, LP-DSA is also very superior to all other methods in $R@k$ on the three test (sub)sets, where LP-ESA holds the second place again. However, differently from $P@k$, the four methods all witness increases in $R@k$ with the rise of k's value on the three (sub)sets, and the increases of LP-DSA and LP-ESA are much more significant than those of ELSA and STPM. Figures 5 (g)-(i) are the results on the F1-score at k ($F@k$) on the three test (sub)sets, where $F@k$ is an integrated metric computed by $R@k$ and $P@k$. Therefore, we have similar observations, revealing the superiority of LP-DSA and LP-ESA.

The results on the truncated mean average precision at k ($MAP@k$) on various test (sub)sets are shown in Figures 5 (j)-(l), where LP-DSA and LP-ESA still dramatically outperform the two baselines at almost all k values. This shows that the recommendation lists generated by LP-DSA and LP-ESA are much better than those of ELSA and STPM, which is a critical improvement for personalized new recommendation, because finding the target news becomes much easier for users.

Evaluation and discussion. In order to evaluate the methods' general performance in news recommendation, according to the values in Figure 5, we calculate the average $P@k$, $R@k$, $F@k$, and $MAP@k$ of these methods on three test (sub)sets, for $1 \leqslant k \leqslant 20$, which are then summarized in Table III.

As shown in Figure 5 and Table III, LP-ESA outperforms ELSA in all cases, e.g., the average $MAP@k$ of LP-ESA is 2.1, 3.9, and 1.5 times better than the one of ELSA on the Old City test subset, the New City test subset, and the whole test set, respectively. These results show the superiority of LP-ESA: they prove the importance of taking into account users' personal interests for location-aware news recommendation and support our assertion that not only geographical contexts, but also personal interests of users should be considered to achieve a satisfactory news recommendation.

Furthermore, LP-ESA also achieves much better news recommendation performance than STPM: the average $P@k$, $R@k$, $F@k$, and $MAP@k$ of LP-ESA on the Old City test subset are 1.7, 7.1, 4.2, and 2.3 times, respectively, as good as those of STPM. The superior performance of LP-ESA may come from the following two aspects: (i) ESA, the topic model used in LP-ESA, which uses Wikipedia for semantic enrichment [34] prior to topic modeling is more effective in geographical topic-based news recommendation than LDA in STMP, especially for short texts on the Web, e.g., tweets. (ii) Instead of using the local history data of users to model localized user preferences directly as in STPM, LP-ESA uses the users' entire history data to model general user profiles, and then uses the local topic distribution to localize the general user profiles for location-aware recommendation; consequently, LP-ESA not only overcomes the cold start problem in STPM (i.e., LP-ESA is also applicable for new locations), but it also achieves a more effective recommendation, as it uses the entire history data to provide more comprehensive information for modeling users' personal interests.

As an improved model of LP-ESA with deep neural networks, the experimental and statistic results in Figure 5 and Table III show that LP-DSA greatly enhances the performance of LP-ESA on all test (sub)sets and in terms of all metrics.
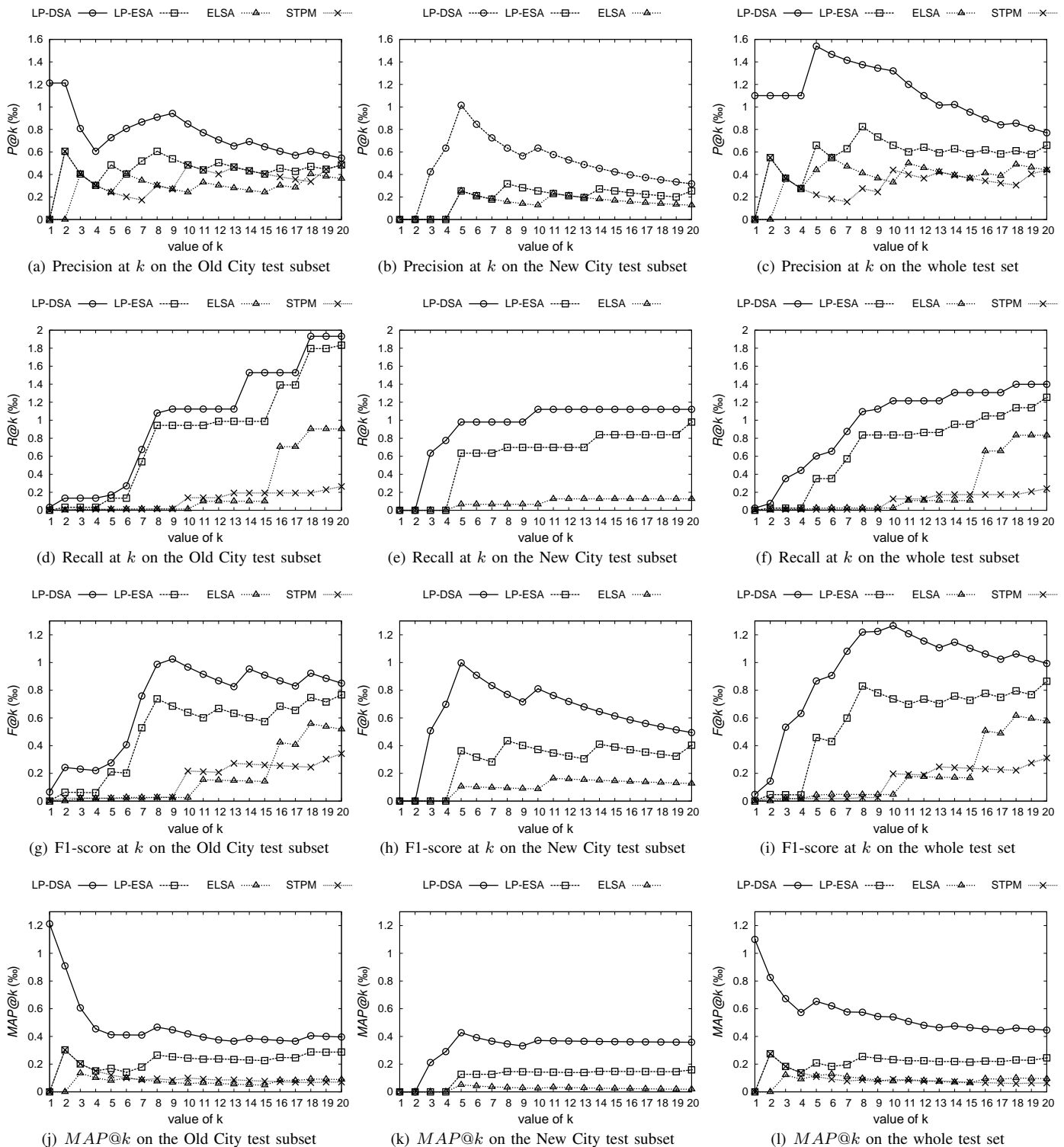
Fig. 5. The news recommendation effectiveness of LP-DSA, LP-ESA, ELSA, and STPM on three test (sub)sets in terms of precision at k ($P@k$), recall at k ($R@k$), F1-score at k ($F@k$), and truncated mean average precision at k ($MAP@k$)

For example, as shown in Table III, the average $P@k$, $R@k$, $F@k$, and $MAP@k$ of LP-DSA are 156.4%, 96.9%, 39.9%, and 172.3%, respectively, better than those of LP-ESA on the whole test set. The improvement is mainly because: (i) By utilizing deep neural networks to extract and compress the incoming Wikipedia-based topic features layer by layer and finally map them to a much denser abstract feature space

with much lower dimensionality, LP-DSA can overcome the huge dimensionality, sparsity, and redundancy problems in LP-ESA. (ii) The deep neural networks in LP-DSA are trained with a recommendation-oriented objective to distinguish the user's local target news from the irrelevant ones. Consequently, the resulting abstract features for user, location, and news profiles are very effective representations for location-aware

TABLE III

AVERAGE $P@k$, $R@k$, $F@k$, AND $MAP@k$ OF DIFFERENT METHODS ON THE THREE TEST (SUB)SETS WHEN $1 \leqslant k \leqslant 20$ (IN ‰)

	Old City test subset				New City test subset				Whole test set			
	P@k	R@k	F@k	MAP@k	P@k	R@k	F@k	MAP@k	P@k	R@k	F@k	MAP@k
STPM	0.357	0.108	0.1504	0.1008	\	\	\	\	0.324	0.098	0.1365	0.0915
ELSA	0.2834	0.2361	0.1687	0.0712	0.1385	0.0836	0.1009	0.0231	0.3773	0.2263	0.1981	0.0843
LP-ESA	0.4443	0.842	0.4919	0.2213	0.1895	0.6053	0.2869	0.1137	0.5675	0.698	0.5776	0.2079
LP-DSA	**0.7656**	**1.0083**	**0.7004**	**0.4792**	**0.486**	**0.932**	**0.6176**	**0.318**	**1.1162**	**0.9768**	**0.9403**	**0.5662**

TABLE IV

TIME COSTS FOR ONLINE RECOMMENDATION (IN MIN)

	ELSA	LP-ESA	LP-DSA
Whole test set	341.3	341.3	**12.98**
Old City test subset	260.8	260.7	**9.781**
New City test subset	75.51	75.62	**2.379**

personalized news recommendation.

In addition, we also note that the average $P@k$, $R@k$, $F@k$, and $MAP@k$ of LP-DSA are 156.4%, 54%, 115.3%, and 179.6% better than those of LP-ESA on the New City test subset, while the corresponding improvements are only 72.3%, 19.8%, 42.4%, and 116.5%, respectively, on the Old City test subset. This observation shows that LP-DSA has more significant improvements to LP-ESA in news recommendation at new locations than those at "old" locations. It may be that, due to the lack of users' history data on new locations, the inferred localized user profiles on the New City test subset are less accurate than those on the Old City test subset; so, LP-ESA is less likely to recommend users' target news to the top positions of recommendation lists at new locations. But by using recommendation-oriented deep neural networks, LP-DSA maps the Wikipedia-based topic space to a more abstract feature space, where the correlations among similar topics (e.g., topics within a category) are strengthened. Consequently, LP-DSA can uncover users' latent preferences to make the inferred abstract localized user profiles more accurate and remedy the problem of lacking history data to a great extent.

E. Efficiency and Scalability.

ESA-based methods, e.g., LP-ESA and ELSA, benefit a lot from using Wikipedia concepts to enrich the semantics; but due to the high volume of concepts, the resulting Wikipedia-based topic space in LP-ESA and ELSA is very large. Consequently, the online news recommendation process that requests to compute the similarities between the localized user and news profiles (resp., the user and news representations) in LP-ESA (resp., ELSA) is usually computationally expensive. As shown in Table IV, even if we have limited the Wikipedia topic space to contain only $8,000$ most frequent topics and speeded up computation using a GPU server, the time costs for the online recommendation processes in LP-ESA using the Old City test subset, the New City test subset, and the whole test set are still up to 260.8, 75.5, and 341.3 minutes, respectively, while those in ELSA are similar. Obviously, such time-consuming recommendation processes are usually unacceptable for the need of real-time response in practice.
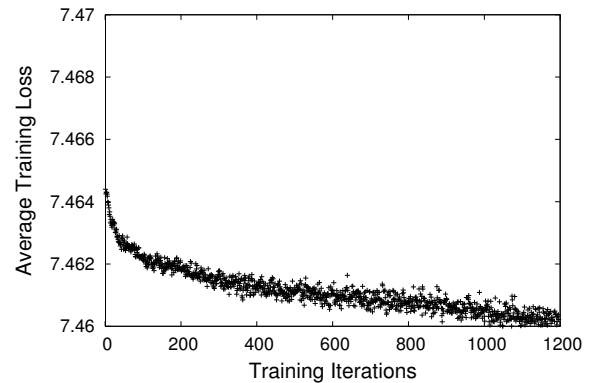

Fig. 6. Average training loss in LP-DSA

Therefore, we propose LP-DSA to solve the high dimensionality problem and enhance the recommendation efficiency and scalability by mapping the Wikipedia topic space to an abstract feature space with lower dimension. As shown in Table IV, although LP-DSA has to pass the general user and news profiles as well as local topic distributions through the whole neural networks prior to computing similarities between abstract localized user and news profiles, the online recommendation process of LP-DSA on the whole test set is still roughly 25 times faster than that of LP-ESA. This proves that, with the help of recommendation-oriented deep neural networks, LP-DSA can achieve much higher recommendation efficiency than LP-ESA; so, it is more scalable.

As for the offline model training of deep neural networks in LP-DSA, the average training loss across all training samples is depicted in Figure 6. In this work, we manually set the number of training iterations to 1200; we do so, because, as shown in Figure 6, although the average loss continuously decreases with the rise of iterations, this improvement is smaller and smaller and becomes relatively stable after the 1000-th iteration. Furthermore, with the help of shared parameters among three neural networks and approximating the news set V with a limited number of news in training, the time needed for each training iteration is greatly reduced. It only takes about 7 hours for LP-DSA to finish 1200 training iterations, showing the good scalability of LP-DSA.

VIII. SUMMARY AND OUTLOOK

In this work, we proposed a location-aware personalized news recommendation with explicit semantic analysis (LP-ESA) which considers both the users' location information and personal interests for news recommendation. Experimental

studies showed that LP-ESA significantly outperforms the state-of-the-art topic-based location-aware news recommendation methods, ELSA and STPM, in terms of $P@k$, $R@k$, $F@k$, and $MAP@k$. However, as LP-ESA is an ESA-based method, each Wikipedia concept is regarded as a topic, and since there are numerous concepts on Wikipedia, the resulting topic space in LP-ESA suffers from the problems of high dimensionality, sparsity, and redundancy, which greatly degrade the performance of LP-ESA. To solve these problems, we further proposed a novel topic model, called deep semantic analysis (DSA), which utilizes deep neural networks to map the (Wikipedia-concept-based) topic space to an abstract, dense, and low dimensional feature space, where the localized similarities between the users and their target news are maximized, and those with irrelevant news are minimized. The resulting news recommendation method is called location-aware personalized news recommendation with deep semantic analysis (LP-DSA). Experimental results showed that LP-DSA further improves the news recommendation performance of LP-ESA: it offers more effective (19.8% to 179.6% better) news recommendation with much lower online recommendation time cost (25 times quicker) than LP-ESA.

In the future, it would be interesting to consider more contextual information (such as timeliness of news and the social relationships of users) to achieve an even better news recommendation. In addition, hybrid learning signals (e.g., combining reconstruction errors with deep-semantic similarities) and more sophisticated neural networks (e.g., convolutional or long short-term memory (LSTM) networks) may be applied to learn a more effective abstract feature space, to further improve the performance of LP-DSA. Finally, going beyond personalized news recommendation, we plan to apply the proposed LP-DSA approach also to other research areas (such as personalized Web search and targeted advertising).

ACKNOWLEDGMENT

REFERENCES

[1] Y. Xie, L. Chen, K. Jia, L. Ji, and J. Wu, "iNewsBox: Modeling and exploiting implicit feedback for building personalized news radio," in *Proceedings of the 22nd ACM International Conference on Information and Knowledge Management*. ACM, 2013, pp. 2485–2488.

[2] F. Garcin, F. Galle, and B. Faltings, "Focal: A personalized mobile news reader," in *Proceedings of the 8th ACM Conference on Recommender systems*. ACM, 2014, pp. 369–370.

[3] K. F. Yeung, Y. Yang, and D. Ndzi, "A proactive personalised mobile recommendation system using analytic hierarchy process and Bayesian network," *Journal of Internet Services and Applications*, vol. 3, no. 2, pp. 195–214, 2012.

[4] F. Abel, Q. Gao, G.-J. Houben, and K. Tao, "Analyzing user modeling on Twitter for personalized news recommendations," in *Proceedings of International Conference on User Modeling, Adaptation, and Personalization*, 2011, pp. 1–12.

[5] J. Bao, M. Mokbel, and C.-Y. Chow, "GeoFeed: A location-aware news feed system," in *Proceedings of the IEEE International Conference on Data Engineering*, 2012, pp. 54–65.

[6] J. Bao and M. F. Mokbel, "GeoRank: An efficient location-aware news feed ranking system," in *Proceedings of the ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 2013, pp. 184–193.

[7] K. V. Qie, "Sensing the news: User experiences when reading locative news." *Future Internet*, vol. 4, pp. 161–178, 2012.

[8] P. M. P. Rosa, J. J. P. C. Rodrigues, and F. Basso, "A weight-aware recommendation algorithm for mobile multimedia systems." *Mobile Information Systems*, vol. 9, no. 2, pp. 139–155, 2013.

[9] Y. Noh, Y.-H. Oh, and S.-B. Park, "A location-based personalized news recommendation," in *Proceedings of the International Conference on Big Data and Smart Computing*, 2014, pp. 99–104.

[10] H. Yin, B. Cui, Y. Sun, Z. Hu, and L. Chen, "LCARS: A spatial item recommender system," *ACM Transactions on Information Systems*, vol. 32, no. 3, pp. 1–37, 2014.

[11] J.-W. Son, A.-Y. Kim, and S.-B. Park, "A location-based news article recommendation with explicit localized semantic analysis," in *Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval.*, 2013, pp. 293–302.

[12] E. Gabrilovich and S. Markovitch, "Wikipedia-based semantic interpretation for natural language processing," *Journal of Artificial Intelligence Research*, vol. 34, no. 1, pp. 443–498, 2009.

[13] Y. Zhang, "GroRec: A group-centric intelligent recommender system integrating social, mobile and Big Data technologies," *IEEE Transactions on Services Computing*, vol. 9, no. 5, pp. 786–795, 2016.

[14] Y. Zhang, M. Chen, S. Mao, L. Hu, and V. C. Leung, "CAP: Community activity prediction based on Big Data analysis," *IEEE Network*, vol. 28, no. 4, pp. 52–57, 2014.

[15] M. Karkali, D. Pontikis, and M. Vazirgiannis, "Match the news: A Firefox extension for real-time news recommendation," in *Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval.* ACM, 2013, pp. 1117–1118.

[16] Y. Mo, J. Chen, X. Xie, C. Luo, and L. T. Yang, "Cloud-based mobile multimedia recommendation system with user behavior information," *IEEE Systems Journal*, vol. 8, no. 1, pp. 184–193, 2014.

[17] W. Xu, C.-Y. Chow, M. L. Yiu, Q. Li, and C. K. Poon, "MobiFeed: A location-aware news feed framework for moving users." *GeoInformatica*, vol. 19, no. 3, pp. 633–669, 2015.

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

[19] Y. Zhang, M. Chen, D. Huang, D. Wu, and Y. Li, "iDoctor: Personalized and professionalized medical recommendations based on hybrid matrix factorization," *Future Generation Computer Systems*, 2016.

[20] Y. Zhou and J. Luo, "Geo-location inference on news articles via multimodal pLSA." in *Proceedings of the ACM International Conference on Multimedia*, 2012, pp. 741–744.

[21] J.-W. Son, Y.-S. Noh, H.-J. Song, and S.-B. Park, "Location comparison through geographical topics," in *Proceedings of the IEEE/WIC/ACM International Joint Conferences on Web Intelligence and Intelligent Agent Technology*, vol. 1, 2012, pp. 311–318.

[22] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 1798–1828, 2013.

[23] A. Van den Oord, S. Dieleman, and B. Schrauwen, "Deep content-based music recommendation," in *Advances in Neural Information Processing Systems*, 2013, pp. 2643–2651.

[24] R. Salakhutdinov, A. Mnih, and G. Hinton, "Restricted Boltzmann machines for collaborative filtering," in *Proceedings of the International Conference on Machine Learning*, 2007, pp. 791–798.

[25] Y. Zuo, J. Zeng, M. Gong, and L. Jiao, "Tag-aware recommender systems based on deep neural networks," *Neurocomputing*, vol. 204, pp. 51–60, 2016.

[26] A. M. Elkahky, Y. Song, and X. He, "A multi-view deep learning approach for cross domain user modeling in recommendation systems," in *Proceedings of the International Conference on World Wide Web*, 2015, pp. 278–288.

[27] L. Page, S. Brin, R. Motwani, and T. Winograd, "The PageRank citation ranking: Bringing order to the Web." 1999.

[28] J. D. Ullman, J. Leskovec, and A. Rajaraman, "Mining of massive datasets," 2011.

[29] Z. Xu, C. Chen, T. Lukasiewicz, Y. Miao, and X. Meng, "Tag-aware personalized recommendation using a deep-semantic similarity model with negative sampling," in *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management.* ACM, 2016, pp. 1921–1924.

[30] L. Bottou, "Stochastic learning," in *Advanced Lectures on Machine Learning*, 2004, pp. 146–168.

[31] A. Richards, "University of Oxford Advanced Research Computing", 2015.

[32] J. Bobadilla, F. Ortega, A. Hernando, and A. Gutiérrez, "Recommender systems survey," *Knowledge-Based Systems*, vol. 46, pp. 109–132, 2013.

[33] F. Aiolli, "Efficient top-n recommendation for very large scale binary rated datasets," in *Proceedings of the ACM Conference on Recommender Systems*, 2013, pp. 273–280.

[34] G. Xu, Z. Wu, G. Li, and E. Chen, "Improving contextual advertising matching by using Wikipedia thesaurus knowledge," *Knowledge and Information Systems*, vol. 43, no. 3, pp. 599–631, 2015.