

Location-Aware News Recommendation Using Deep Localized Semantic Analysis

Cheng Chen^{1,2}, Thomas Lukasiewicz³, Xiangwu Meng^{1,2,*}, and Zhenghua Xu³

¹ Beijing Key Laboratory of Intelligent Telecommunications Software and Multimedia, Beijing University of Posts and Telecommunications, China

² School of Computer Science, Beijing University of Posts and Telecommunications, China
{ccbupt, mengxw}@bupt.edu.cn

³ Department of Computer Science, University of Oxford, UK
firstname.lastname@cs.ox.ac.uk

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 strongly correlated with their geographical contexts. Consequently, many research efforts have been put on location-aware news recommendation; the explored approaches can mainly be divided into physical distance-based and geographical topic-based ones. As for geographical topic-based location-aware news recommendation, ELSA is the state-of-the-art geographical topic model: it has been reported to outperform many other topic models, e.g., BOW, LDA, and ESA. However, the Wikipedia-based topic space in ELSA suffers from the problems of high dimensionality, sparsity, and redundancy, which greatly degrade the recommendation performance of ELSA. Therefore, to overcome these problems, in this work, we propose three novel geographical topic feature models, CLSA, ALSA, and DLSA, which integrate clustering, autoencoders, and recommendation-oriented deep neural networks, respectively, with ELSA to extract dense, abstract, low dimensional, and effective topic features from the Wikipedia-based topic space for the representation of news and locations. Experimental results show that (i) CLSA, ALSA, and DLSA all greatly outperform the state-of-the-art geographical topic model, ELSA, in location-aware news recommendation in terms of both the recommendation effectiveness and efficiency; (ii) Deep Localized Semantic Analysis (DLSA) achieves the most significant improvements: its precision, recall, MRR, and MAP are all about 3 times better than those of ELSA; while its recommendation time-cost is only about 1/29 of that of ELSA; and (iii) DLSA, ALSA, and CLSA can also remedy the “cold-start” problem by uncovering users’ latent news preferences at new locations.

Keywords: Location-aware news recommendation, explicit semantic analysis, autoencoders, deep neural networks.

1 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,

*Corresponding author.

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 [1,8,11,22].

In typical news recommendation systems, a user's news preferences are usually learned from his/her news reading history or other online activity history; so his/her news preferences are (almost) static in these systems. However, in real-world contexts, users' news preferences usually evolve with the change of their locations; e.g., 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 research directions: physical distance-based and geographical topic-based approaches.

Specifically, physical distance-based news recommendation [2,3,14,16] 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 their locations 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 [12,19,23] are proposed to achieve a more generic location-aware news recommendation, where, 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 geographical topic-based location-aware news recommendation, and a range of topic models (such as Latent Dirichlet Allocation (LDA) [5], Explicit Semantic Analysis (ESA) [10], Probabilistic Latent Semantic Analysis (PLSA) [24], and their improved models [12,19,23]) have been used.

The state-of-the-art geographical topic model is *Explicit Localized Semantic Analysis (ELSA)* [19], which is reported to outperform many other geographical topic models (e.g., BOW, LDA, and ESA) in geographical topic-based location-aware news recommendation. 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) [10], where Wikipedia concepts are regarded as topics; consequently, by considering link information between the corresponding concepts of local topics, both locations (e.g., country, city, or venue) and news are represented as topic vectors (called *localized location profiles* and *localized news profiles*, respectively) and the relevance score between a user and a candidate news article is estimated by the similarity between the corresponding localized location and news profiles; finally, the news with top-k highest relevance scores are recommended to the user.

However, since the volume of concepts is enormous (millions) on Wikipedia, the resulting Wikipedia-based topic space in ELSA is very high dimensional. Consequently,

the process of online news recommendation in ELSA is very time-consuming, which is unacceptable for the need of real-time online responses in practice. In addition, the Wikipedia-based topic space also suffers from the problems of sparsity and redundancy, which degrade the news recommendation effectiveness of ELSA to a great extent.

Therefore, to achieve better recommendation performance, in this work, we first propose two geographical topic feature models, *Clustering-based Localized Semantic Analysis (CLSA)* and *Autoencoder-based Localized Semantic Analysis (ALSA)* to address these problems by *topic feature modeling*. Generally, CLSA and ALSA integrate clustering and autoencoders (neural networks), respectively, with ELSA to extract denser, more abstract and lower dimensional topic features from the Wikipedia-based topic space in ELSA for the representations of news and locations. Our experimental studies show that CLSA and ALSA both improve the performance of ELSA in location-aware news recommendations. However, these two solutions still suffer from the following drawback: the learning objectives of clustering in CLSA and autoencoders in ALSA are to minimize the within-cluster distances and the reconstruction errors, respectively, which are not directly correlated to the objective of news recommendation, i.e., distinguishing the user’s local target news from the irrelevant ones; so the resulting cluster-based or autoencoder-based topic feature representations of news and locations may not be very effective in news recommendation.

Motivated by this observation, we further propose another novel geographical topic feature model, called *Deep Localized Semantic Analysis (DLSA)* model, to address the high dimensionality, sparsity, and redundancy problems in ELSA. DLSA utilizes deep neural networks to map the Wikipedia-based topic space in ELSA to an abstract, dense, and low dimensional topic feature space, where the localized similarities between the locations and users’ local target news are maximized, and those with the users’ irrelevant news are minimized. DLSA has the following advantage: the deep neural networks in DLSA are trained with a recommendation-oriented learning objective, i.e., to differentiate the users’ local target news from the irrelevant ones, so the resulting deep topic feature representations of news and locations are more effective for location-aware news recommendations than ALSA and CLSA. Consequently, the performance of DLSA is superior to those of CLSA and ALSA in location-aware news recommendation. Although we only investigate their applications in ELSA, the proposed models, CLSA, ALSA, and DLSA, can easily be used to tackle similar problems in other topic models.

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

- We identify the high dimensionality, sparsity, and redundancy problems in the Wikipedia-based topic space of ELSA, which greatly degrades ELSA’s location-aware news recommendation performance.
- We thus propose three novel geographical topic feature models (CLSA, ALSA, and DLSA) to address these problems by topic feature modeling. These three models integrate clustering, autoencoders, and recommendation-oriented deep neural networks, respectively, with ELSA to obtain an abstract, dense, low dimensional, and effective topic feature representation for locations and news.
- Extensive experiments are conducted using a public real-world dataset. The results show that: (i) the proposed CLSA, ALSA, and DLSA all greatly outperform the state-of-the-art geographic topic model, ELSA, in location-aware news recommendation in both the recommendation effectiveness and efficiency; (ii) DLSA achieves

the most significant improvements: its precision, recall, MRR, and MAP are all about 3 times better than those of ELSA, while its recommendation time-cost is only about 1/29 of that of ELSA; (iii) DLSA, ALSA, and CLSA can remedy “cold-start” problems by uncovering users’ latent news preferences at new locations.

2 Related Work

2.1 News Recommendation

Typical news recommendation systems aim to recommend to users the news that match their personal interests best [8,11]. Users’ interests in news are usually modeled by their explicit ratings or browsing histories (e.g., visited pages, reading times, and downloads). Both heuristic [1] and model-based methods [22] are proposed for news recommendations: 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. [1] 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. [22] used Bayesian networks to predict levels of interesting news categories for users and to provide real-time personalized news recommendation.

2.2 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 happening nearby than those far away from them. Therefore, more and more research efforts have been put into location-aware news recommendations, which mainly focus on two research directions: physical distance-based and geographical topic-based.

As for physical distance-based news recommendation, GeoFeed [2] and GeoRank [3] recommend to users some news happening at the users’ current locations or within a given range, where GeoRank uses only static location points of both users and news, while GeoFeed allows news with spatial extent; Pedro et al. [16] utilized the Euclidean distance between the locations of users and news articles to measure the importance of news articles; LocaNews [14] keeps three versions of news and offers to users the most suitable ones according to their different distances to the locations of news; Wen et al. [21] proposed a news stream recommendation framework, called MobiFeed, to further investigate news recommendation based on users’ moving tracks.

However, in the real-world context, the descriptions 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. Therefore, the application of physical distance-based methods is limited.

As for more generic location-aware news recommendations, 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 geographical topic-based location-aware news recommendation, and a range of topic models (such as Latent Dirichlet Allocation (LDA) [5], Explicit Semantic Analysis (ESA) [10], Probabilistic Latent Semantic Analysis (PLSA) [24], and their improved models [12,19,23]) have been used. The state-of-the-art geographical topic model in topic-based location-aware news recommendation is *Explicit Localized Semantic Analysis (ELSA)* [19], which is reported to outperform many other topic models, e.g., BOW, LDA [5], and ESA [10]. However, the Wikipedia-based topic space in ELSA suffers from the problems of high dimensionality, sparsity, and redundancy, which greatly degrade the recommendation performance of ELSA. Therefore, in this work, we propose three novel geographical topic feature models (CLSA, ALSA, and DLSA) to overcome these problems by topic feature modeling and to achieve better recommendation performance.

2.3 Recommendation Using Deep Learning

Due to its capability to extract effective representations [4], deep learning has already been successfully applied in many online recommendation applications, such as music recommendation [20], movie recommendation [17], tag-aware recommendation [25], and multi-view item recommendation [9].

Similarly to our work, the recommendation system proposed in [9] is also based on deep neural networks with a recommendation-oriented training objective. But [9] is very different from the DLSA proposed here: (i) [9] is not a location-aware model, so its recommendation is not sensitive to the changes of users' geographical contexts; (ii) [9] does not aim to solve the huge dimensionality, sparsity, and redundancy problems in the Wikipedia-based topic space; and (iii) [9] has to train own parameters for each neural network; while, in our work, shared parameters are applied in deep neural networks, as 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.

3 Explicit Localized Semantic Analysis in News Recommendation

Explicit localized semantic analysis (ELSA) [19] is the state-of-the-art topic model in 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 briefly review ELSA in this section.

ELSA is an ESA-based [10] solution, where each Wikipedia concept is considered as a potential topic, and each location and news article is represented as a Wikipedia-based topic vector. Fig. 1 shows the overall process of ELSA. First of all, ELSA collects for each location a set of documents with the corresponding geo-tags (denoted as D_l) as the description of this location. Then, these geo-tagged documents and the candidate set of news articles V are mapped onto a Wikipedia-based topic space (denoted as Z) to generate for each location or news article a topic vector, which is represented as a probability distribution over topics, called *general location profile* (φ_l) or *general news profile* (φ_v), respectively. Consequently, the topics related to a location are the ones with non-zero probability values in this location's general profile. Since these topics are generally dependent on one another, ELSA further uses the link information within the corresponding Wikipedia concepts to construct a topic dependency graph and then applies PageRank [13] to estimate a *local topic distribution* (θ_l). With the help of the local

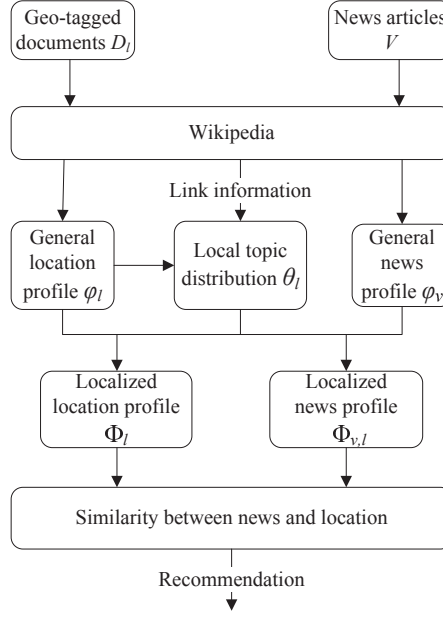


Fig. 1. Overall process of ELSA

topic distribution, the location, and the news, general profiles are localized to obtain the local topic representations of locations and news, which are called *localized location profile* (Φ_l) and *localized news profile* ($\Phi_{v,l}$). Finally, ELSA estimates the similarities between news and locations based on their localized profiles, and makes recommendations by offering to users the news articles with top- k similarity scores to their current locations. For the detailed inferences of φ_l , φ_v , θ_l , Φ_l , and $\Phi_{v,l}$, please refer to [19].

4 Topic Feature Modeling

Although ELSA benefits from using Wikipedia-concept-based topics for semantic enrichment, due to the huge volume of concepts (millions) on Wikipedia, the resulting topic space in ELSA suffers from the problems of high dimensionality, sparsity, and redundancy, which greatly degrade ELSA’s recommendation effectiveness and efficiency.

Therefore, we propose to apply *topic feature modeling* to address these problems and to achieve a better performance in location-aware news recommendation. Generally, the process of topic feature modeling takes the general location profile (φ_l), general news profile (φ_v), and local topic distribution (θ_l) in ELSA as inputs, and exploits either clustering or deep learning techniques to extract dense, abstract, low dimensional, and effective topic features from the Wikipedia-based topic space for the representations of news and locations. The solutions that utilize clustering, autoencoders, and recommendation-oriented deep neural networks for topic feature modeling are presented in the rest of this section.

4.1 Clustering-based Localized Semantic Analysis

Due to its capability in extracting abstract and low dimensional features [18], in this work, we adopt hierarchical clustering as the first solution for the topic feature modelling in ELSA; the resulting geographical topic feature model that integrates clustering with ELSA is called *Clustering-based Localized Semantic Analysis (CLSA)*.

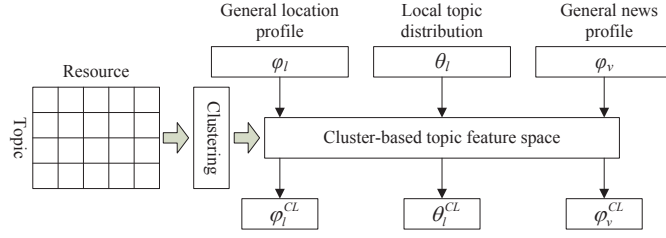


Fig. 2. Clustering-based topic feature modeling

As shown in Fig. 2, the clustering-based topic feature modeling in CLSA first represents each Wikipedia topic $z \in Z$ as a vector of weights over the set of resources, i.e., the locations and news articles, where the weight on each dimension is measured by the probability of the resource generated from the corresponding topic. Then, hierarchical clustering [18] groups all topics into a number of clusters based on the distances between their corresponding resource vectors. Finally, with the resulting well-learned clusters, CLSA converts the Wikipedia-based topic space to a cluster-based topic feature space; so the representations of general location profile ϕ_l , local topic distribution θ_l , and general news profile ϕ_v are converted from Wikipedia-based topic vectors to cluster-based topic feature vectors, denoted ϕ_l^{CL} , θ_l^{CL} , and ϕ_v^{CL} , respectively.

CLSA tackles the high dimensionality, sparsity, and redundancy problems in ELSA, because: (i) the cluster-based topic feature space is lower dimensional than the Wikipedia-based topic space in ELSA; (ii) each cluster contains several topics, so sparsity is diminished; and (iii) redundant topics are aggregated to a cluster to reduce redundancy.

As for location-aware recommendation, similarly to ELSA, given ϕ_l^{CL} , θ_l^{CL} , and ϕ_v^{CL} , CLSA first obtains the cluster-based localized location and news profiles (denoted Φ_l^{CL} and $\Phi_{v,l}^{CL}$, respectively), which are formally defined as

$$\Phi_l^{CL} = (\phi_l^{CL})^T \cdot \theta_l^{CL}, \quad \Phi_{v,l}^{CL} = (\phi_v^{CL})^T \cdot \theta_l^{CL}. \quad (1)$$

Then, given a user at a location l , CLSA generates news recommendations by ranking all news $v \in V$ according to their relevance to l (denoted $R_{l,v}^{CL}$), where the relevance is estimated by the cosine similarity between Φ_l^{CL} and $\Phi_{v,l}^{CL}$. Formally,

$$R_{l,v}^{CL} = \text{Sim}(\Phi_l^{CL}, \Phi_{v,l}^{CL}) = \frac{\Phi_l^{CL} \cdot \Phi_{v,l}^{CL}}{\|\Phi_l^{CL}\| \cdot \|\Phi_{v,l}^{CL}\|}. \quad (2)$$

4.2 Autoencoder-based Localized Semantic Analysis

Besides clustering, autoencoders are another method to model low dimensional, dense, and abstract representations of raw data [25]. So, in this work, we also employ autoencoders as another solution for the high dimensionality, sparsity, and redundancy problems in ELSA; the resulting geographical topic model that integrates autoencoders with ELSA is called *Autoencoder-based Localized Semantic Analysis (ALSA)*.

Autoencoders are neural networks consisting of two parts: an encoder and a decoder. As shown in Fig. 3, to conduct topic feature modeling, autoencoders first take ϕ_l , θ_l , and ϕ_v as inputs, which are passed through multiple hidden layers in encoders. The intermediate outputs $f_i(l)$, $f_i(\theta)$, and $f_i(v)$ of the i -th hidden layers for location, local topics distribution, and news can be formally defined as

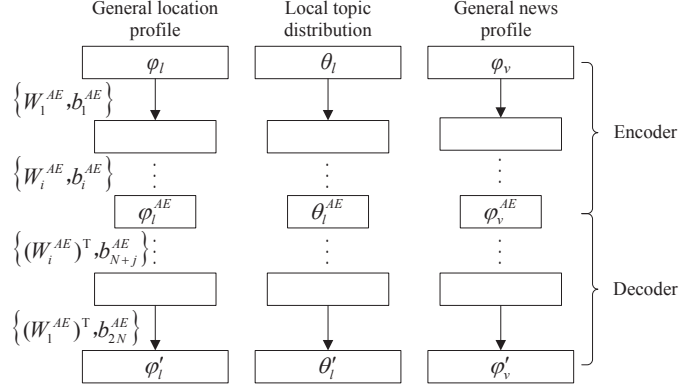


Fig. 3. Autoencoder-based topic feature modeling

$$f_i(l) = \sigma(W_i^{AE} f_{i-1}(l) + b_i^{AE}), \quad (3)$$

$$f_i(\theta) = \sigma(W_i^{AE} f_{i-1}(\theta) + b_i^{AE}), \quad (4)$$

$$f_i(v) = \sigma(W_i^{AE} f_{i-1}(v) + b_i^{AE}), \quad (5)$$

where $i = 1, \dots, N$; N is the total number of hidden layers in the encoder (decoder); W_i^{AE} and b_i^{AE} are the weight matrix and bias vector for the i -th hidden layer; $\sigma(\cdot)$ is the sigmoid activation function; and $f_0(l) = \phi_l$, $f_0(\theta) = \theta_l$, $f_0(v) = \phi_v$.

The outputs of the final layers of encoders ($f_N(l)$, $f_N(\theta)$, and $f_N(v)$) are the autoencoder-based topic feature representations for location (denoted ϕ_l^{AE}), local topic distribution (denoted θ_l^{AE}), and news (denoted ϕ_v^{AE}), respectively. Formally

$$\phi_l^{AE} = f_N(l), \quad \theta_l^{AE} = f_N(\theta), \quad \phi_v^{AE} = f_N(v).$$

Furthermore, the decoders in ALSA take ϕ_l^{AE} , θ_l^{AE} , and ϕ_v^{AE} as inputs and pass them through another N layers. Since we use tied-weights autoencoders, the weight matrices in the decoder are the transposes of those in the encoder. Formally, we have

$$f_{N+j}(l) = \sigma((W_{N-(j-1)}^{AE})^T f_{N+(j-1)}(l) + b_{N+j}^{AE}), \quad (6)$$

$$f_{N+j}(\theta) = \sigma((W_{N-(j-1)}^{AE})^T f_{N+(j-1)}(\theta) + b_{N+j}^{AE}), \quad (7)$$

$$f_{N+j}(v) = \sigma((W_{N-(j-1)}^{AE})^T f_{N+(j-1)}(v) + b_{N+j}^{AE}), \quad (8)$$

where $j = 1, \dots, N$. The outputs of the decoder are the reconstructed general location profile, the reconstructed local topic distribution, and the reconstructed general news profile, denoted ϕ'_l , θ'_l , and ϕ'_v , respectively. Formally,

$$\phi'_l = f_{2N}(l), \quad \theta'_l = f_{2N}(\theta), \quad \phi'_v = f_{2N}(v).$$

As for the training of autoencoders, the learning objective of autoencoders in ALSA is to minimize the differences between input and reconstructed data, called *reconstruction errors*. Therefore, the loss function of ALSA is as follows:

$$L^{AE}(\Theta) = \frac{1}{2} \sum_{(l,v)} (\|\phi'_l - \phi_l\| + \|\theta'_l - \theta_l\| + \|\phi'_v - \phi_v\|), \quad (9)$$

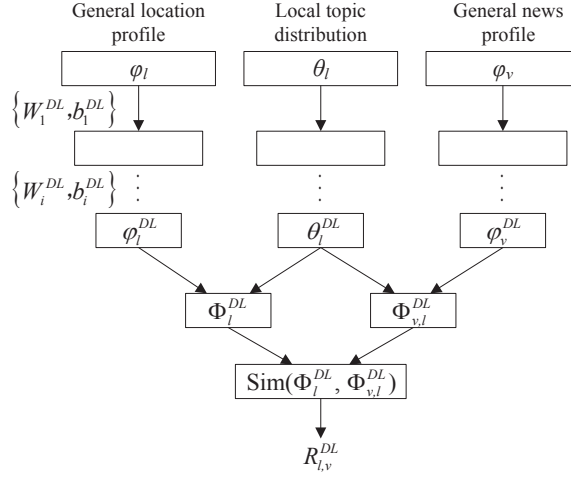


Fig. 4. Topic feature modeling based on recommendation-oriented deep neural networks

where Θ represents the set of parameters $\{W_i^{AE}, b_j^{AE}\}$ ($i = 1, \dots, N; j = 1, \dots, 2N$) in autoencoders.

After training, given the well-modeled autoencoder-based topic feature representations φ_l^{AE} , θ_l^{AE} , and φ_v^{AE} , ALSA first generates the autoencoder-based localized location and news profiles by

$$\Phi_l^{AE} = (\varphi_l^{AE})^T \cdot \theta_l^{AE}, \quad (10)$$

$$\Phi_{v,l}^{AE} = (\varphi_v^{AE})^T \cdot \theta_l^{AE}. \quad (11)$$

Then, given a user at a location l , ALSA generates the location-aware recommendations based on the relevance of all news $v \in V$ to l , which is computed by the cosine similarity between Φ_l^{AE} and $\Phi_{v,l}^{AE}$. Formally,

$$R_{l,v}^{AE} = \text{Sim}(\Phi_l^{AE}, \Phi_{v,l}^{AE}) = \frac{\Phi_l^{AE} \cdot \Phi_{v,l}^{AE}}{\|\Phi_l^{AE}\| \cdot \|\Phi_{v,l}^{AE}\|}. \quad (12)$$

4.3 Deep Localized Semantic Analysis

However, the learning objectives of clustering in CLSA and autoencoders in ALSA are to minimize the within-cluster distances and minimize the reconstruction errors, respectively, which are not directly correlated to the objective of the news recommendation, i.e., distinguishing the users' local target news from the irrelevant ones; so the resulting clustering-based and autoencoder-based topic feature representations of news and locations may not be very effective in news recommendation.

Therefore, we further propose another novel geographical topic feature model, called *Deep Localized Semantic Analysis (DLSA)*, to address the high dimensionality, sparsity, and redundancy problems in ELSA. DLSA also applies deep neural networks for topic feature modeling; however, instead of using autoencoders, DLSA newly integrates ELSA with recommendation-oriented deep neural networks, which maps the Wikipedia-based topic space to an abstract, dense, and low dimensional topic feature space, where the localized similarities between the locations and the users' local target (resp., irrelevant) news are maximized (resp., minimized). Since the deep neural networks in DLSA are trained with a recommendation-oriented learning objective, i.e., to differentiate the users' local target news from the irrelevant ones, the resulting deep

topic feature representations of news and locations in DLSA are more effective for location-aware news recommendation than CLSA and ALSA.

As shown in Fig. 4, similarly to the encoders in ALSA, the three deep neural networks in DLSA also take φ_l , θ_l , and φ_v as inputs, and the intermediate outputs h_i of the i -th hidden layers are formally defined as follows:

$$h_i(l) = \tan(W_i^{DL} h_{i-1}(l) + b_i^{DL}), \quad (13)$$

$$h_i(\theta) = \tan(W_i^{DL} h_{i-1}(\theta) + b_i^{DL}), \quad (14)$$

$$h_i(v) = \tan(W_i^{DL} h_{i-1}(v) + b_i^{DL}), \quad (15)$$

where $i = 1, \dots, N$; \tan is used as the activation function; and $h_0(l) = \varphi_l$, $h_0(\theta) = \theta_l$, $h_0(v) = \varphi_v$. Furthermore, the intermediate outputs in the N -th hidden layers are the deep topic feature representations for the general location profile (denoted φ_l^{DL}), local topics distribution (denoted θ_l^{DL}), and general news profile (denoted φ_v^{DL}); formally,

$$\varphi_l^{DL} = h_N(l), \quad \theta_l^{DL} = h_N(\theta), \quad \varphi_v^{DL} = h_N(v).$$

Given φ_l^{DL} , θ_l^{DL} , and φ_v^{DL} , the deep localized location and news profiles are defined as

$$\Phi_l^{DL} = (\varphi_l^{DL})^T \cdot \theta_l^{DL}, \quad \Phi_{v,l}^{DL} = (\varphi_v^{DL})^T \cdot \theta_l^{DL}. \quad (16)$$

Then, for a user at a location l , the similarity between l and a news article v is measured via the cosine similarity between their deep localized profiles at l (i.e., Φ_l^{DL} and $\Phi_{v,l}^{DL}$):

$$\text{Sim}(\Phi_l^{DL}, \Phi_{v,l}^{DL}) = \frac{\Phi_l^{DL} \cdot \Phi_{v,l}^{DL}}{\|\Phi_l^{DL}\| \cdot \|\Phi_{v,l}^{DL}\|}. \quad (17)$$

Differently from CLSA and ALSA, instead of using cosine similarity directly, the relevance scores of news v to given users at locations l are measured by applying the softmax function on the resulting similarities of all news at l , which are then used to generate location-aware recommendation lists. Formally,

$$R_{l,v}^{DL} = e^{\text{Sim}(\Phi_l^{DL}, \Phi_{v,l}^{DL})} / \sum_{v' \in V} e^{\text{Sim}(\Phi_l^{DL}, \Phi_{v',l}^{DL})}. \quad (18)$$

Intuitively, to achieve good location-aware news recommendations, the local target news should have higher relevance scores than irrelevant ones. We thus conduct the model training in DLSA with a recommendation-oriented objective to maximize the relevance scores of local target news; equivalently, we maximize the localized similarities between locations and their local target news and minimize those with irrelevant ones. Formally, it is equivalent to minimize the following loss function:

$$\begin{aligned} L^{DL}(\Theta) &= -\sum_{(l,v^*)} \log(R_{l,v^*}^{DL}) \\ &= -\sum_{(l,v^*)} [\log(e^{\text{Sim}(\Phi_l^{DL}, \Phi_{v^*,l}^{DL})}) - \log(\sum_{v' \in V} e^{\text{Sim}(\Phi_l^{DL}, \Phi_{v',l}^{DL})})], \end{aligned} \quad (19)$$

where Θ is the set of parameters $\{W_i^{DL}, b_i^{DL}\}$ ($i = 1, \dots, N$) in DLSA; tuple (l, v^*) is a training sample, indicating that v^* is a local target news to the user at location l .

As for the training of DLSA (resp., ALSA), we first initialize the weight matrices W_i^{DL} (resp., W_i^{AE}) using the random normal distribution and initialize the bias vectors b_i^{DL} (resp., b_i^{AE}) to be zero vectors; the model is then trained via stochastic gradient descent [7], which is a gradient-based optimization algorithm; finally, the training

Table 1. Statistic information of the dataset

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

Table 2. Details of training set and test (sub)sets

	Users	News	Locations	Samples
Training set	1,558	51,399	1,089	86,086
Test set	1077	11,965	1805	12,235
Old City test subset	941	10,734	581	11,000
New City test subset	527	1,231	1,224	1,235

stops when the model converges or reaches the maximum training iterations. Two optimization solutions are used to enhance the training efficiency of DLSA and ensure its scalability in practice: (i) Rather than training own parameters for each neural network, networks in DLSA share parameters; and (ii) negative sampling is used to further reduce the training cost of DLSA. As shown in our prior work [26], sharing parameters is reasonable and negative sampling can greatly enhance the model’s training efficiency by hundreds of times while maintaining almost the same training effectiveness.

In summary, applying deep neural networks for topic feature modeling in DLSA and ALSA is capable to overcome the high dimensionality, sparsity, and redundancy problems in ELSA, because (i) the number of nodes in the hidden layer is much smaller than that in the input layer, so the dimensionality of the resulting deep (or autoencoder-based) topic feature space is much lower than that of the Wikipedia-based topic space in ELSA; and (ii) deep neural networks in DLSA and ALSA extract more abstract and denser features layer-by-layer, so sparsity and redundancy problems are addressed.

5 Experiments

We evaluate the performances of ELSA, CLSA, ALSA, and DLSA using a publicly available Twitter dataset [1], which consists of 2, 316, 204 tweets posted by 1, 619 users ($|U| = 1, 619$). About half 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 tool¹ 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 1.

We assume that a user is specified only by his/her location. Then, if a user posts a tweet containing an URL to a news article v^* with a city name (location) l in its content, the user is believed to be interested in v^* at l , from which a sample (l, v^*) is generated, indicating v^* is a local target news to location l . Consequently, a total of 98, 321 samples are obtained from the dataset; we randomly select 85% of the samples as the training set and the remaining 15% as the test set. To evaluate the different recommendation performance on “old” locations (“old” cities), which have appeared in the training set, and on new locations (new cities), which do not exist in the training set, we further divide the test set into two subsets: for each sample (l, v^*) in the test set, if the training

¹OpenCalais at <https://opencalais.com/>

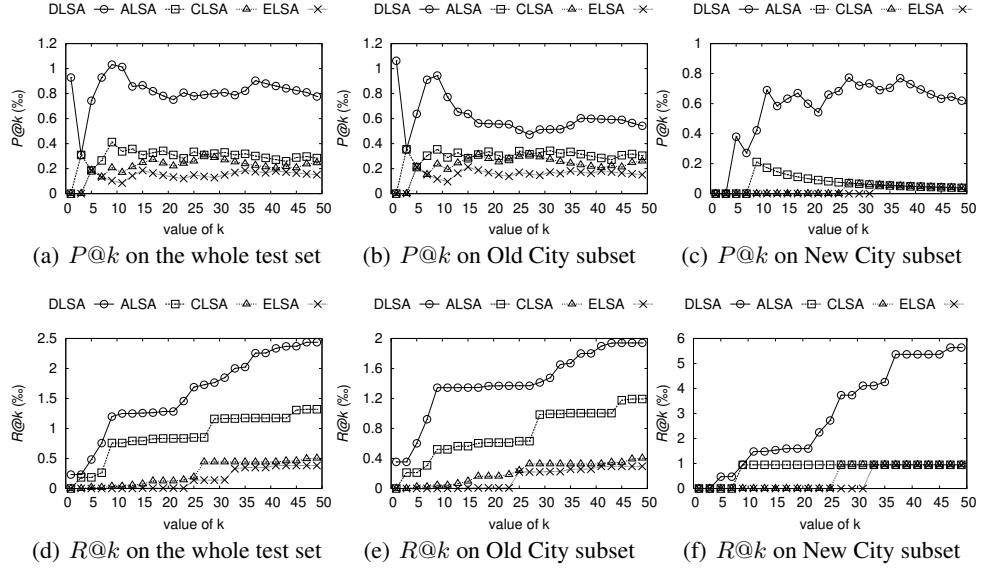


Fig. 5. Performance of DLSA, ALSA, CLSA and ELSA in terms of $P@k$ and $R@k$

set also contains some samples related to location l , l is seen as an “old” city (location), so (l, v^*) is added to Old City test subset; otherwise, l is a new city (location), and the sample (l, v^*) is added to New City test subset. The details are summarized in Table 2.

Finally, a Wikipedia snapshot of August 11, 2014 is used for the Wikipedia-based semantic enrichment [10] in ELSA, CLSA, ALSA, and DLSA, resulting in 1,301,900 concepts with 1,618,970 distinct terms. To cut down the calculation and memory cost, we select 8,000 most frequent concepts as the Wikipedia-based topic space Z .

All methods are implemented using Python and Theano and run on a server of Oxford University’s ARC facility [15] with an NVIDIA Tesla K40 GPU and 12GB GPU memory. ELSA is implemented based on [19]; # of clusters in CLSA is empirically set to 1024; and the parameters of ALSA and DLSA are empirically set as follows: (i) # of hidden layers in DLSA and in the encoder of ALSA: $N = 3$; (ii) # of neurons in the 1st, 2nd, and 3rd hidden layer: 1024, 512, and 256, respectively; (iii) learning rate for model training: 0.0001; additionally, (iv) ALSA adds another two hidden layers for the decoder, and # of neurons in the 4th and 5th layer are 512 and 1024, respectively.

The most popular metrics for the evaluation of recommendation systems are precision and recall [6]. Since 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* ($P@k$) and *recall at k* ($R@k$). Since users always prefer to have their target news ranked in the front of the recommendation list, we also use *mean average precision* (MAP) and *mean reciprocal rank* (MRR) as evaluation metrics, which give greater importance to news ranked higher.

5.1 Main Results

Fig. 5 depicts the news recommendation performance of DLSA, ALSA, CLSA, and ELSA on three test (sub)sets in terms of precision at k ($P@k$) and recall at k ($R@k$), where k varies from 1 to 50. In addition, Table 3 shows the performance of DLSA, ALSA, CLSA, and ELSA on three test (sub)sets in terms of MRR and MAP.

Table 3. Performance of DLSA, ALSA, CLSA, and ELSA in terms of MRR and MAP

	The whole test set		Old City test subset		New City test subset	
	MRR(%)	MAP(%)	MRR(%)	MAP(%)	MRR(%)	MAP(%)
ELSA	0.1243	0.2483	0.1301	0.2799	0.0726	0.0951
CLSA	0.1734	0.3530	0.1796	0.3714	0.1185	0.1434
ALSA	0.2192	0.4514	0.2174	0.4441	0.2356	0.2532
DLSA	0.4737	0.8491	0.4001	0.9132	1.130	0.7217

Generally, as shown in both Fig. 5 and Table 3, the proposed three geographical topic feature models, DLSA, ALSA, and CLSA, all greatly outperform the Wikipedia-based topic model, ELSA, in location-aware news recommendation in terms of all evaluation metrics on all three test (sub)sets. This finding demonstrates that applying clustering or deep learning techniques for topic feature modeling can address the high dimensionality, sparsity, and redundancy problems in ELSA and greatly enhances the location-aware news recommendation effectiveness.

Furthermore, with the help of the recommendation-oriented deep neural networks, DLSA achieves a much better recommendation performance than ALSA and CLSA. For example, on the whole test set, the precision, recall, MRR, and MAP of DLSA are all roughly 3 times better than those of ELSA, about 2 times better than those of CLSA, and about double of ALSA. The superior performance of DLSA is mainly because DLSA uses a recommendation-oriented learning objective, which is directly correlated with distinguishing the user’s local target news from the irrelevant ones; so, the resulting deep topic features for the representations of news and location profiles in DLSA are much more effective for news recommendations than those generated by clustering and autoencoders, whose learning objectives are not directly related to recommendations.

We also note that DLSA, ALSA, and CLSA generally achieve more significant improvements to ELSA at new locations than those at “old” locations. For example, the MAP’s of DLSA, ALSA, and CLSA are 63.2%, 224.5%, and 14.6 times, respectively, better than that of ELSA on the New City test subset, while the corresponding improvements are only 38.0%, 67.1%, and 3.08 times on the Old City test subset. It may be that due to the lack of users’ history data on new locations, the inferred localized location profiles on the New City test subset are less accurate than those on the Old City test subset, so ELSA is less likely to recommend users’ local target news to the top positions of recommendation lists at new locations. But, by using clustering or deep learning techniques for topic feature modeling, the proposed geographical topic feature models, CLSA, ALSA, and DLSA, map the Wikipedia-based topic space to a more abstract feature topic space, where the correlations among similar topics (e.g., topics within a category) are strengthened. Thus, CLSA, ALSA, and DLSA can uncover users’ latent localized news preferences to make the inferred abstract localized location profiles more accurate. So, DLSA, ALSA, and CLSA also remedy the “cold-start” problem.

5.2 Efficiency and Scalability

News recommendation requests real-time responses in practice; so the recommendation efficiency is crucial for online location-aware news recommendation. However, due to the high volume of concepts in Wikipedia, the Wikipedia-based topic space in ELSA is very huge. Consequently, the online news recommendation process that requests to compute the similarities between the localized location and news profiles in ELSA is usually computationally expensive. As shown in Table 4, even if we have limited the

Table 4. Total time-costs for online recommendations (in min)

	ELSA	CLSA	ALSA	DLSA
The whole test set	160.4	18.59	5.552	5.699
Old City test subset	142.1	16.89	5.006	5.160
New City test subset	15.67	3.997	0.5961	0.5906

Wikipedia topic space to contain only 8,000 most frequent topics and speeded up computation using a GPU server, the total (resp., average) time-costs for the online recommendation processes in ELSA are still up to 142.1, 15.67, and 160.4 (resp., 0.0129, 0.0127, and 0.0131) minutes on Old City, New City, and the whole test (sub)sets, respectively, which are usually unacceptable in practice.

Therefore, to ensure scalability in the real-world context, the proposed DLSA (resp., CLSA and ALSA) solve the high dimensionality problem by mapping the Wikipedia-based topic space to a deep (resp., clustering- and autoencoder-based) topic feature space with much lower dimensionality. As shown in Table 4, although DLSA and ALSA have to pass data through the well-trained neural networks prior to compute similarities between deep or autoencoder-based localized location and news profiles, the total online recommendation time-costs of DLSA and ALSA on the whole test set are still only about 1/29 of that of ELSA. In addition, although the total time-cost of CLSA is higher than those of DLSA and ALSA, it is still much lower than that of ELSA. Overall, these findings prove that, with the help of topic feature modeling, the proposed geographical topic feature models, CLSA, ALSA, and DLSA, can achieve much higher recommendation efficiency than ELSA; so, it is more scalable in practice.

6 Summary and Outlook

In this work, we proposed three novel geographical topic feature models, CLSA, ALSA, and DLSA, which integrate clustering, autoencoders, and recommendation-oriented deep neural networks, respectively, with ELSA to address the high dimensionality, sparsity, and redundancy problems existing in ELSA’s Wikipedia-based topic space. Experimental studies showed that CLSA, ALSA, and DLSA all significantly outperform the state-of-the-art geographical topic model, ELSA, in location-aware news recommendation in terms of both effectiveness and efficiency, while DLSA achieves the best performance: it offers more effective (about 3 times better) location-aware news recommendation with much lower online recommendation time cost (about 28 times quicker) than ELSA. In addition, DLSA, ALSA, and CLSA can also remedy the “cold-start” problem by uncovering users’ latent news preferences at new locations.

In the future, it would be interesting to consider user personal preferences and more contextual information, such as timeliness of news and the social relationships of users, to achieve better personalized context-aware 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) neural networks) may be applied to learn a more effective abstract topic feature space, and so to further improve the performance of DLSA.

Acknowledgments. This work is supported by the Mutual Project of Beijing Municipal Education Commission. Thomas Lukasiewicz and Zhenghua Xu are supported by the UK EPSRC Grants EP/J008346/1, EP/L012138/1, and EP/M025268/1, and by The Alan Turing Institute under the EPSRC Grant EP/N510129/1.

References

1. F. Abel, Q. Gao, G.-J. Houben, and K. Tao. Analyzing user modeling on Twitter for personalized news recommendations. In *Proc. of UMAP*, pp. 1–12, 2011.
2. J. Bao, M. Mokbel, and C.-Y. Chow. GeoFeed: A location-aware news feed system. In *Proc. of ICDE*, pp. 54–65, 2012.
3. J. Bao and M. F. Mokbel. GeoRank: An efficient location-aware news feed ranking system. In *Proc. of SIGSPATIAL*, pp. 184–193, 2013.
4. Y. Bengio, A. Courville, and P. Vincent. Representation learning: A review and new perspectives. *TPAMI*, 35(8):1798–1828, 2013.
5. D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent Dirichlet allocation. *JMLR*, 3:993–1022, 2003.
6. J. Bobadilla, F. Ortega, A. Hernando, and A. Gutiérrez. Recommender systems survey. *Knowledge-Based Systems*, 46:109–132, 2013.
7. L. Bottou. Stochastic learning. In *Proc. of ALML*, pp. 146–168, 2003.
8. W. Chu and S.-T. Park. Personalized recommendation on dynamic content using predictive bilinear models. In *Proc. of WWW*, pp. 691–700, 2009.
9. A. M. Elkahky, Y. Song, and X. He. A multi-view deep learning approach for cross domain user modeling in recommendation systems. In *Proc. of WWW*, pp. 278–288, 2015.
10. E. Gabrilovich and S. Markovitch. Wikipedia-based semantic interpretation for natural language processing. *JAIR*, 34(1):443–498, 2009.
11. L. Li, W. Chu, J. Langford, and R. E. Schapire. A contextual-bandit approach to personalized news article recommendation. In *Proc. of WWW*, pp. 661–670, 2010.
12. Y. Noh, Y.-H. Oh, and S.-B. Park. A location-based personalized news recommendation. In *Proc. of BigComp*, pp. 99–104, 2014.
13. L. Page, S. Brin, R. Motwani, and T. Winograd. The PageRank citation ranking: Bringing order to the Web. 1999.
14. K. V. Qie. Sensing the news: User experiences when reading locative news. *Future Internet*, 4:161–178, 2012.
15. A. Richards. University of Oxford Advanced Research Computing, 2015.
16. P. M. P. Rosa, J. J. P. C. Rodrigues, and F. Basso. A weight-aware recommendation algorithm for mobile multimedia systems. *Mobile Info. Sys.*, 9(2):139–155, 2013.
17. R. Salakhutdinov, A. Mnih, and G. Hinton. Restricted Boltzmann machines for collaborative filtering. In *Proc. of ICML*, pp. 791–798, 2007.
18. A. Shepitsen, J. Gemmell, B. Mobasher, and R. Burke. Personalized recommendation in social tagging systems using hierarchical clustering. In *Proc. of RecSys*, pp. 259–266, 2008.
19. J.-W. Son, A.-Y. Kim, and S.-B. Park. A location-based news article recommendation with explicit localized semantic analysis. In *Proc. of SIGIR*, pp. 293–302, 2013.
20. A. Van den Oord, S. Dieleman, and B. Schrauwen. Deep content-based music recommendation. In *Proc. of NIPS*, pp. 2643–2651, 2013.
21. 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*, 19(3):633–669, 2015.
22. K. F. Yeung and Y. Yang. A proactive personalized mobile news recommendation system. In *Proc. of DeSE*, pp. 207–212, 2010.
23. H. Yin, B. Cui, Y. Sun, Z. Hu, and L. Chen. LCARS: A spatial item recommender system. *ACM TOIS*, 32(3):1–37, 2014.
24. Y. Zhou and J. Luo. Geo-location inference on news articles via multimodal pLSA. In *Proc. of ACM MM*, pp. 741–744, 2012.
25. Y. Zuo, J. Zeng, M. Gong, and L. Jiao. Tag-aware recommender systems based on deep neural networks. *Neurocomputing*, 204:51–60, 2016.
26. 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 *Proc. of CIKM*, pp. 1921–1924, 2016.