Geotagging Social Media Posts to Landmarks Using Hierarchical BERT (Student Abstract)

Menglin Li and Kwan Hui Lim

Information Systems Technology and Design Pillar
Singapore University of Technology and Design, Singapore, Singapore
menglin_li@mymail.sutd.edu.sg, kwanhui_lim@sutd.edu.sg

Abstract

Geographical information provided in social media data is useful for many valuable applications. However, only a small proportion of social media posts are explicitly geotagged with their posting locations, which makes the pursuit of these applications challenging. Motivated by this, we propose a 2-level hierarchical classification method that builds upon a BERT model, coupled with textual information and temporal context, which we denote HierBERT. As far as we are aware, this work is the first to utilize a 2-level hierarchical classification approach alongside BERT and temporal information for geolocation prediction. Experimental results based on two social media datasets show that HierBERT outperforms various state-of-art baselines in terms of accuracy and distance error metrics.

Introduction

With the rapid emergence of social media, social data with geotags (locations) open up new possibilities for many significant applications, such as location-based services ranging from targeted advertising to crisis detection (George et al. 2021). However, the low ratio of geotagged social data makes the pursuit of the aforementioned applications challenging. Therefore, addressing the issue of geotagging social content turns out to be extremely important, especially regarding locations with fine granularity, e.g., at the landmark level.

The information contained in social media data can be divided into three types: social content, social network, and social context. Thanks to the rich textual information contained in social data, it is becoming a common approach to perform geolocation prediction using text classification methods. Compared with typical text classification tasks, geotagging social posts is much more challenging due to the characteristics of social data. The posting text could be very noisy, containing acronyms, misspellings, special tokens, among others. Furthermore, the social context could involve textual information, temporal information, spatial information, and even image information synchronously. The popularity of different landmarks could vary dramatically, which also reflects on the number of social media posts for each location. This leads to the imbalance issue when inferring the real-time posting locations, thus increasing the challenge in social media post geolocation. However, many related works did not consider this issue (Ouaret et al. 2019).

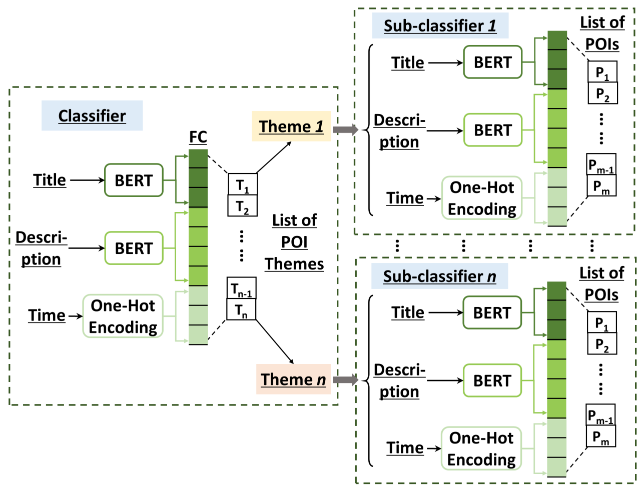

Figure 1: An Illustration of HierBERT using textual information and temporal information.

In this paper, we utilize the idea of hierarchical classification to solve the problem of geolocation prediction to further improve prediction performance. A 2-level classification model, HierBERT, is proposed based on BERT with textual and temporal information. We conduct experiments on two datasets (Flickr and Twitter) and results show that HierBERT out-performs various state-of-the-art baselines.

Network Structure

As shown in Figure 1, the 2-level model consists of one level-1 classifier and n level-2 sub-classifiers. For each classifier or sub-classifier, the inputs contain textual information and temporal information. Here we use BERT (Devlin et al. 2019) to generate the semantic representations of words, which shows excellent performance in multiple NLP tasks. These word representations of each textual inputs are then concatenated with the one-hot encoded vectors of the temporal inputs. Finally, they are passed to a softmax layer to produce the probabilities of each label. The level-2 sub-classifiers are slightly different from the level-1 classifier. Since level-2 labels are sub-classes of level-1 labels, the inputs for each level-2 sub-classifier are the corresponding sub-sets of level-1 inputs.

Inputs Representations

The textual inputs involve the title and description of the social media post. To utilize them, we perform tokenization first, to transform each sentence into several tokens. A special token [CLS] is also inserted at the beginning of the sentence, which would be used for classification directly later. BERT would then convert each token into three embeddings: token embedding, segment embedding, and position embedding. Token embedding is the 1-dimension vector of a token. For segment embedding, it is used to distinguish the first sentence and the second sentence since BERT could take in sentence pairs as inputs. As there are no sentence pairs in our datasets, all segment embeddings are equal to 0. Position embedding denotes the position of each token in the sentence. Thereafter, these three embeddings are passed into BERT to generate the semantic representations of words. BERT is used to get general representations of words, which contain semantic information and rich contextual information, from large-scale unlabelled language materials. Two pre-trained tasks are involved in BERT: Masked Language Model (MLM) and Next Sentence Prediction to generate semantic representations. For time-related features, we extract temporal elements from date and time, like hour, weekday, month, and so on, then represent them as one-hot encoding and use them for classification directly.

Hierarchical Structure

We utilize hierarchical classification to address the geolocation prediction problem. The implementation of hierarchical classification requires the support of structured label data, which is satisfied by the structure of Points-of-Interest (POI) data. Each POI has multi-level labels, such as POI themes, POI sub-themes. In our 2-level model, we use POI themes as level-1 labels and POIs as level-2 labels. As displayed in Figure 1, the level-1 classifier outputs a list of POI themes. Take the theme T_1 for example. T_1 denotes the POI theme: leisure/recreation. Then a sub-classifier is constructed for this theme and the output is a list of POIs under this theme, including Carlton Gardens South, Batman Park, and so on.

Experiments

Dataset and Baselines

Our social media datasets are collected from the Twitter (266k tweets) and Flickr (202k photos) platforms, then mapped to POIs based on proximity. We compared our model against several state-of-art geotagging baselines models, including MNB-Ngrams (Chong and Lim 2018), CNN-PreW (Kim 2014), CNN-1Hot (Johnson and Zhang 2014), MLP-BoW (Rahimi, Cohn, and Baldwin 2017), CNN-TextTime (Lim et al. 2019), and 1-level BERT (Devlin et al. 2019).

Methods	Flickr		Twitter	
	$Acc@5$	$MDist$	$Acc@5$	$MDist$
HierBERT	**0.9119**	**264.2**	**0.9050**	**265.8**
1-level BERT	*0.7847*	*403.5*	*0.7994*	*350.1*
CNN-TextTime	0.8245	283.3	0.8069	351.9
CNN-PreW	0.6933	533.0	0.7727	408.5
CNN-1Hot	0.7003	679.3	0.7689	362.0
MLP-BoW	0.7070	656.6	0.7635	408.1
MNB-3grams	0.6737	631.5	0.7360	500.7

Table 1: Evaluation results

Results

We use Accuracy@5 ($Acc@5$) and Mean Distance Error ($MDist$) to evaluate the geotagging performance. Table 1 shows that HierBERT provides the best performance against baselines by between 10.6% to 35.3% in terms of $Acc@5$ and 6.7% to 61.1% in terms of $MDist$.

Conclusion

In this paper, we employ BERT to generate semantic representations and incorporate them with temporal information. To address the geolocation problem, we propose a 2-level hierarchical classification method HierBERT, which builds upon BERT. The experiments based on Flickr and Twitter datasets both demonstrate the effectiveness of HierBERT against various baselines.

Acknowledgments

This research is funded in part by SUTD under grant SRG-ISTD-2018-140.

References

Chong, W.-H.; and Lim, E.-P. 2018. Exploiting user and venue characteristics for fine-grained tweet geolocation. *ACM TOIS*, 36(3): 1–34.

Devlin, J.; Chang, M.-W.; Lee, K.; and Toutanova, K. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proc. of NAACL*, 4171–4186.

George, Y.; Karunasekera, S.; Harwood, A.; and Lim, K. H. 2021. Real-time Spatio-temporal Event Detection on Geotagged Social Media. *Journal of Big Data*, 8(91): 1–28.

Johnson, R.; and Zhang, T. 2014. Effective use of word order for text categorization with convolutional neural networks. *arXiv preprint arXiv:1412.1058*.

Kim, Y. 2014. Convolutional Neural Networks for Sentence Classification. In *Proc. of EMNLP'14*, 1746–1751.

Lim, K. H.; Karunasekera, S.; Harwood, A.; and George, Y. 2019. Geotagging tweets to landmarks using convolutional neural networks with text and posting time. In *IUI'19*.

Ouaret, R.; Birregah, B.; Soulier, E.; Auclair, S.; and Boulahya, F. 2019. Random Forest location prediction from social networks during disaster events. In *SNAMS'19*.

Rahimi, A.; Cohn, T.; and Baldwin, T. 2017. A Neural Model for User Geolocation and Lexical Dialectology. In *Proc. of ACL'17*, 209–216.