Hashtag2Vec: Learning Hashtag Representation with Relational Hierarchical Embedding Model

Jie Liu, Zhicheng He and **Yalou Huang**
College of Computer and Control Engineering, Nankai University, Tianjin, China.
jliu@nankai.edu.cn, hezhicheng@mail.nankai.edu.cn, huangyl@nankai.edu.cn

Abstract

Hashtags have always been important elements in many social network platforms. Semantic understanding of hashtags is a critical and fundamental task for many applications on social networks, such as event analysis, theme discovery, information retrieval, etc. However, this task is challenging due to the sparsity, polysemy, and synonymy of hashtags. In this paper, we investigate the problem of hashtag embedding by combining the short text content with the various heterogeneous relations in social networks. Specifically, we first establish a network with hashtags as its nodes. Hierarchically, each of the hashtag nodes is associated with a set of tweets and each tweet contains a set of words. Then we devise an embedding model, called Hashtag2Vec, which exploits multiple relations of hashtag-hashtag, hashtag-tweet, tweet-word, and word-word relations based on the hierarchical heterogeneous network. In addition to embedding the hashtags, our proposed framework is capable of embedding the short social texts as well. Extensive experiments are conducted on two real-world datasets, and the results demonstrate the effectiveness of the proposed method.

1 Introduction

Hashtags play an important role in information diffusion in many social network platforms by organizing messages and highlighting topics. Taking Twitter as an example, there are about 240 million active users who tweet more than 500 million tweets every day, and a quarter of the tweets are tagged with hashtags [Shi *et al.*, 2016]. Hashtags are keyword-based tags, describing the content of a tweet, e.g., #superbowl, #nfl, etc. Hashtags can be used for various purposes, including brand promotion [Adamopoulos and Todri, 2015], micro-meme discussions [Huang *et al.*, 2010], and tweets categorization [Godin *et al.*, 2013]. Moreover, as the number of tweets is overwhelmingly large, hashtags can also be used to facilitate information retrieval [Efron, 2010], which makes tweets easily searchable and more accessible. Therefore, knowledge discovery of hashtags is of great importance to enable many applications such as targeted recommendations, content organization, and event analysis.

Despite its value and significance, learning meaningful and effective representations for hashtags and their associated texts (tweets) remains in its infancy due to the following challenges: 1) The uncontrolled creation and adoption of hashtags lead to the problems of data sparsity, polysemy, and synonymy. 2) The structural relations, such as co-occurrence of hashtags and hashtag sharing of tweets, reflect crucial semantic information, but how to model the heterogeneous relations is a non-trivial task. 3) Other than the structural relation information, the content information also plays an important role in the semantic modeling of hashtags. However, the nature of short text produces very sparse bag-of-words representation and constrains the following representation learning. In summary, to better learn hashtag representations, it is highly desirable to develop techniques that comprehensively consider the heterogeneous information and jointly learn the representations of different objects.

To tackle these problems, we investigate the hashtag embedding problem and propose a hierarchical embedding framework with heterogeneous relations, which is named as Hashtag2Vec. We first establish a hashtag network according to the co-occurrence relation. Two nodes (hashtags) are connected if they co-occur in some tweet. Each hashtag has two-level hierarchical text information, corresponding to tweets and words respectively. Comparing with existing Network Embedding (NE) models [Perozzi *et al.*, 2014; Yang *et al.*, 2015], the network considered in our task contains multiple kinds of objects and more complex structures. Hence it depicts the hashtag-centered social texts more comprehensively. To cope with the hierarchical heterogeneous network embedding, we devise embedding models for different relations, which jointly factorize structure matrices and content matrices. The structure matrices consist of the hashtag-hashtag co-occurrence matrix and the hashtag-tweet interaction matrix. The content matrices are the tweet-word matrix and the word-word matrix. Utilizing tweets as side information, we can harvest the hashtag co-occurrence relationship. With the equivalent matrix factorization form of DeepWalk [Perozzi *et al.*, 2014], the distributed vector representations of hashtags can be obtained. And owing to the factorization form of DeepWalk and word embedding [Levy and Goldberg, 2014], the multiple matrices are factorized within a unified framework simultaneously. Comparing with other NE approaches, our model produces two important byproducts, the tweets' and the words' embeddings. Thus, the representation learning of each kind of object, i.e., hashtag, tweet, word, can be mutually enhanced. Although Twitter is a representative social media platforms for the proposed approach, it can also be applied to other social media platforms, such as Facebook, Flickr, etc. Experiments on two real-world datasets demonstrate that our proposed approaches achieve superior performance against other state-of-the-art methods.

Our contributions are summarized as follows:

- We propose a hierarchical embedding approach for hashtags and tweets as well.
- We further propose to incorporate multiple kinds of heterogeneous information from the perspectives of both content and structure to enhance the representation learning.
- We extensively evaluate our proposed approach on both hashtag and tweet clustering tasks on two real-world datasets. And the experimental results show significant improvement.

2 Related Work

There has been extensive research conducted to discover meaningful information and topics from the massive, noisy, and short texts generated by social network users. Hashtags added by users can aid the semantic understanding of the short social texts. [Sedhai and Sun, 2015] created a dataset named HSpam14 for hashtag-oriented spam filtering in tweets. A learning-to-rank approach for modeling hashtag relevance is proposed in [Shi *et al.*, 2016] to address real-time recommendation of hashtags to streaming news articles. [Vicient and Moreno, 2015] used nouns in Wikipedia categories to link hashtags to WordNet concepts, and proposed a domain-independent method to discover topics in Twitter. Traditional topic models can also be applied to short text understanding. Two PLSA-style topic models are introduced in [Ma *et al.*, 2014] to capture the implicit relations between latent topics in tweets and their corresponding hashtags. The PLSA-style models also verify the impact of social factors on hashtag annotation by the introduction of social network regularization. [Weston *et al.*, 2014] described a convolutional neural network (CNN) that learns feature representations for short textual posts using hashtags as supervised signals. An attention-based CNN model is developed to exploit word semantics to model the hashtag recommendation task as a semantic classification problem [Gong and Zhang, 2016].

There is limited work explicitly exploit the relational information of hashtags and tweets. [Wang *et al.*, 2016] developed an LDA based model which incorporates the hashtag graph constructed from their co-occurrences. While in this paper, we propose an NE based hashtag representation learning model. Some embedding approaches have already been proposed for general network data including DeepWalk [Perozzi *et al.*, 2014], LINE [Tang *et al.*, 2015], and node2vec [Grover and Leskovec, 2016], which learn the embeddings of vertices with the neighborhood information. Max-margin DeepWalk (MMDW) [Tu *et al.*, 2016] is proposed to learn discriminative network representations by utilizing labels of vertices. To consider the heterogeneous content information accompanied with the vertices, text-associated DeepWalk (TADW) [Yang *et al.*, 2015] is proposed to improve DeepWalk with text information. And CANE models the semantic relationships between vertices more precisely through learning context-aware embeddings for vertices with mutual attention mechanism [Tu *et al.*, 2017]. Being different from the existing NE models, our approach deals with more complex networks with multiple heterogeneous nodes and relations, and it also considers the content information.

3 Approach

In this work, we propose to learn the representations of different types of objects via a joint embedding framework, Hashtag2Vec. Under the proposed framework, the representation U^h for hashtags, U^t for tweets, and U^w for words can be learned simultaneously and reinforced mutually. As shown in Fig. 1, the considered hierarchical heterogeneous graph $G = (V^h \cup V^t \cup V^w, E^{hh} \cup E^{ht} \cup E^{tw} \cup E^{ww})$, has three kinds of vertices: hashtags V^h, tweets V^t and words V^w; and four kinds of edges: hashtag-hashtag E^{hh}, hashtag-tweet E^{ht}, tweet-word E^{tw} and word-word E^{ww}. Given the heterogeneous graph G, each type of relationship can be represented as an adjacency matrix, namely, M^{hh}, M^{ht}, M^{tw}, and M^{ww}.

3.1 Content-based Embedding

The semantic meanings of hashtags are conveyed by their associated short texts, i.e., tweets. In this work, we first propose a hierarchical content based embedding approach. It is designed to capture the semantic information about the hashtag-tweet-word hierarchical content, as shown in Fig. 1. It takes the advantages of both document representation learning and word distributed representation learning.

Tweet Level Embedding

Tweets can be naturally seen as documents consisting of words. Standard topic modeling approaches, like NMF, LDA, can be applied to tweet topic discovery. Here, we exploit a neural embedding model. Specifically, to model the proximity of a tweet and a word in the embedding space, we define their joint probability as:

$$p^{tw}(i,j) = \sigma(u_i^t \cdot u_j^w) = \frac{1}{1+\exp(-u_i^t \cdot u_j^w)}, \tag{1}$$

where $u_i^t \in \mathbb{R}^K$ and $u_j^w \in \mathbb{R}^K$ are K dimensional embedding vectors of the i-th tweet and the j-th word respectively. And the logistic function $\sigma(\cdot,\cdot)$ is adopted to transform representation similarity into co-occurrence probability. Eq. (1) defines a distribution $p^{tw}(\cdot,\cdot)$ over tweet and word pairs, and its empirical distribution $\hat{p}^{tw}(\cdot,\cdot)$ can be obtained from the adjacent matrix M^{tw}. Here we define it as the normalized adjacency weight $\hat{p}^{tw}(i,j) = m_{ij}^{tw} / \sum_{(i,j') \in E^{tw}} m_{ij'}^{tw}$, where m_{ij}^{tw} is an entry of M^{tw}. To approximate the proximity information in embedding space, we can minimize the distance between these two distributions:

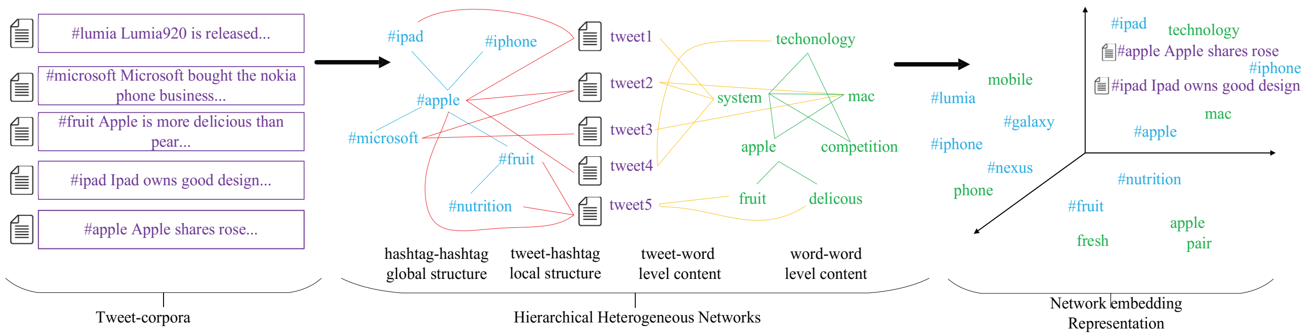

#lumia Lumia920 is released...
#microsoft Microsoft bought the nokia phone business...
#fruit Apple is more delicious than pear...
#ipad Ipad owns good design...
#apple Apple shares rose...
Tweet-corpora
#ipad
#iphone
#apple
#microsoft
#fruit
#nutrition
tweet1
tweet2
tweet3
tweet4
tweet5
techonology
system
mac
apple
competition
fruit
delicous
hashtag-hashtag global structure
tweet-hashtag local structure
tweet-word level content
word-word level content
Hierarchical Heterogeneous Networks
#ipad
technology
#apple Apple shares rose
#iphone
#ipad Ipad owns good design
mobile
#lumia
#galaxy
mac
#apple
#iphone
#nexus
phone
#nutrition
#fruit
apple
fresh
pair
Network embedding Representation

Figure 1: The illustration of hierarchical heterogeneous network embedding for hashtag representation learning.

$$L^{tw} = d(\hat{p}^{tw}(\cdot,\cdot), p^{tw}(\cdot,\cdot)), \tag{2}$$

where $d(\cdot,\cdot)$ measures the dissimilarity between the two distributions, and we choose the squared Euclidean distance. Thus the loss function L^{tw} can be rewritten as:

$$L^{tw} = \sum_{(i,j)\in E^{tw}} (\hat{p}^{tw}(i,j) - p^{tw}(i,j))^2. \tag{3}$$

Word Level Embedding

Compared with regular documents, tweets are featured with its shortness. To deal with this challenge, we further introduce the word-word relation to capture the word co-occurrences in local contexts. The objective is to predict context words, which are surrounding words within a fixed window, with the given current word. For the same reason as in tweet-word content embedding, we use a similar loss function.

$$L^{ww} = \sum_{(i,j)\in E^{ww}} (\hat{p}^{ww}(i,j) - p^{ww}(i,j))^2, \tag{4}$$

where $p^{ww}(i,j) = \sigma(u_i^w \cdot u_j^w)$. To be precise, the empirical probability $\hat{p}^{ww}(i,j)$ is the pointwise mutual information (PMI) [Church and Hanks, 1989] of the respective word and context pairs, shifted by a global constant. With the word-word adjacency matrix M^{ww}, PMI between a word i and its context word j is defined as:

$$\hat{p}^{ww}(i,j) = \text{PMI}(i,j) = \log \frac{m_{ij}^{ww} \cdot |D|}{\#(i) \cdot \#(j)}. \tag{5}$$

where $\#(i) = \sum_{(i,j')\in E^{ww}} m_{ij'}^{ww}$, $\#(j) = \sum_{(i',j)\in E^{ww}} m_{i'j}^{ww}$, and $D = \sum_{i'} \#(i')$ that summarizes over all possible word-word pairs. Due to PMI matrix tends to be ill-posed and dense, Shifted Positive PMI (SPPMI) matrix is considered as a better alternative to PMI matrix [Levy and Goldberg, 2014]. Hence we have

$$\hat{p}^{ww}(i,j) = \text{SPPMI}(i,j) = \max\{\text{PMI}(i,j) - \log k,\ 0\}. \tag{6}$$

3.2 Structure-based Embedding

In addition to the content, the structure of hashtag network also conveys meaningful information of hashtags and tweets. The structural information can be captured from two perspectives. On the one hand, hashtags co-occurred should be embedded as similar representation vectors. On the other hand, distributed representations of hashtags and the tweets which they occur in should be similar, too. Therefore, we propose to encode the structural information by combining the merits from these two aspects.

Global Structure-based Embedding

The network of hashtags is established by their co-occurrence relation E^{hh}, which is a global structure of hashtags. The adjacency weight M^{hh} is calculated by the number of co-occurrences. DeepWalk [Perozzi *et al.*, 2014] is effective in embedding nodes of a network, while it is not capable of learning representation for heterogeneous networks. Here we first adapt DeepWalk for hashtag network without considering other types of objects. Specifically, the proximity in embedding space of two hashtags can be captured via the following joint probability function:

$$p^{hh}(i,j) = \sigma(u_i^h \cdot u_j^h), \tag{7}$$

where $u_i^h \in \mathbb{R}^K$ is the low-dimensional vector representation of the i-th hashtag. Given the adjacency matrix M^{hh}, the empirical distribution $\hat{p}^{hh}(\cdot,\cdot)$ is defined as the logarithm of the average probability that hashtag i randomly walks to hashtag j in t steps:

$$\hat{p}^{hh}(i,j) = \log([e_i(M^{hh} + (M^{hh})^2 + \cdots + (M^{hh})^t)]_j / t), \tag{8}$$

where e_i is the one-hot vector with the i-th element equals 1. The objective is to minimize the distance between these two distributions:

$$L^{hh} = \sum_{(i,j)\in E^{hh}} (\hat{p}^{hh}(i,j) - p^{hh}(i,j))^2. \tag{9}$$

Local Structure-based Embedding

Another important relation is the local interaction between hashtags and tweets. It is intuitive that the topic of a hashtag is discussed by the tagged tweets. Hence, the tweets should have similar topics with the adopted hashtags. Compared with the hashtag-hashtag relation, the hashtag-tweet relation contributes to the learning of hashtag embedding from another perspective. Similar to Eq. (7), we utilize a joint probability function to convey the co-occurrence information,

$$p^{ht}(i,j) = \sigma(u_i^h \cdot u_j^t). \tag{10}$$

The empirical distribution $\hat{p}^{ht}(\cdot,\cdot)$ can also be obtained from adjacency matrix M^{ht} like in Eq. (8). As the hashtag-tweet relation is enough to depict their semantic proximity, we directly take the normalized adjacency weight as the empirical probability $\hat{p}^{ht}(i,j)$:

$$\hat{p}^{ht}(i,j) = \frac{m_{ij}^{ht}}{\sum_{(i,j')\in E^{ht}} m_{ij'}^{ht}}. \tag{11}$$

Then a loss function that minimizes the distance between $p^{ht}(\cdot,\cdot)$ and $\hat{p}^{ht}(\cdot,\cdot)$ can be derived:

$$L^{ht} = \sum_{(i,j)\in E^{ht}} \left(\hat{p}^{ht}(i,j) - p^{ht}(i,j)\right)^2. \tag{12}$$

3.3 Heterogeneous Joint Embedding

To learn the embedding of the heterogeneous network, we embed the four networks by jointly minimizing the following objective function:

$$J = \min_{\theta} L^{tw} + L^{ww} + L^{hh} + L^{ht} + \lambda\Omega(\theta) \tag{13}$$

where θ is the set of parameters, $\theta = \{U^h, U^t, U^w\}$, $\Omega(\cdot)$ is the regularization term $||U^h||_F^2 + ||U^t||_F^2 + ||U^w||_F^2$, and λ is a hyper-parameter.

Due to the shortness of tweets, M^{tw} can be very sparse. Hence we aggregate the tweets with same hashtags into pseudo-documents, which are more dense, to create the hashtag-word matrix M^{hw}. So, we have a loss function for the hashtag-word relation:

$$L^{hw} = \sum_{(i,j)\in E^{hw}} (\hat{p}^{hw}(i,j) - p^{hw}(i,j))^2, \tag{14}$$

where$p^{hw}(i,j) = \sigma(u_i^h \cdot u_j^w)$ and $\hat{p}^{hw}(i,j) = m_{ij}^{hw} / \sum_{(i,j')\in E^{hw}} m_{ij'}^{hw}$. The joint embedding objective function with aggregation is:

$$J^{agg} = \min_{\theta} L^{hw} + L^{ww} + L^{hh} + L^{ht} + \lambda\Omega(\theta). \tag{15}$$

We evaluate both Hashtag2Vec presented in Eq. (13) and Hashtag2Vec(agg) presented in Eq. (15) in the Experiments section.

The optimization problems in Eq. (13) and Eq. (15) can be solved by using any gradient descent method. In this paper, we adopt the widely applied stochastic gradient descent (SGD) method.

4 Experiments

4.1 Data

In order to verify the effectiveness of our model, we use two tweet collections, Tweet2011 and Tweet2015. Tweet2011 is a tweet collection published in TREC 2011 microblog track. Tweet2015 [Wang *et al.*, 2016] is a collection that crawled from Twitter.com in the period from June 17th to June 23rd, 2015 by selecting hot keywords. The raw datasets are preprocessed with stemming and retweets removing. The statistics of the resulted datasets are shown in Table 1. As Hashtag2Vec is an unsupervised model, the hashtag network constructed from the whole dataset is used for the learning of the embeddings.

4.2 Evaluation Metric

Clustering hashtags and tweets are key problems in targeted recommendation, content organization, event detection and analysis. Hence, we conduct hashtag and tweet clustering to evaluate the effectiveness of the representations produced by the comparison methods. After learning the distributed representation of hashtags, the similarity between hashtags can be calculated in a semantic space. Our evaluations are based on H-Score [Yan *et al.*, 2013] which is a commonly used clustering metric. H-score reflects the ratio of average intra-cluster distance and the average inter-cluster distance. Smaller H-score indicates better performance.

4.3 Comparison Models

We compare our method with many state-of-the-art methods for learning embeddings of hashtags and tweets. The comparison models can be classified into three categories, including content-only models, structure-only models and structure-and-content models. The content-only models are NMF [Lee and Seung, 2000], TWTM [Li *et al.*, 2013b], TWDA [Li *et al.*, 2013a], ATM [Rosen-Zvi *et al.*, 2010], and HGTM [Wang *et al.*, 2016]. As NMF can not learn the content based representations of hashtag and tweet simultaneously, hashtag embeddings are learned from the hashtag-word matrix M^{hw}, while tweet embeddings are learned from M^{tw}. The structure-only models are LINE [Tang *et al.*, 2015], DeepWalk [Perozzi *et al.*, 2014] and node2vec [Grover and Leskovec, 2016]. The structure-and-content models are TADW [Yang *et al.*, 2015] and CANE [Tu *et al.*, 2017].

4.4 Clustering Hashtags and Tweets

To evaluate the capacity of Hashtag2Vec in heterogeneous network embedding, we conduct hashtag and tweet clustering experiments. Table 2 and Table 3 shows the hashtag and tweet clustering performance of each comparison method on Tweet2011 and Tweet2015 respectively. From the tables, we have the following observations:

- As NMF learns hashtag embedding from pesudo-documents, it performs much better than other content-based models in hashtag clustering, while the performance is contrary in tweet clustering. This indicates that the sparsity of tweet constrains the classical topic modeling approaches, such as NMF. The models, like TWTM and TWDA, are designed for short text modeling, and they can achieve much better performance than NMF in tweet clustering task.
- Among the content-based models, ATM and HGTM, which consider the hashtag relation to some extent, outperform other content-based models.
- The content-based models are generally better than the structure-only NE models, which implies that using structure information only is limited in hashtag embedding task. Compared with the sparse hashtag network, the texts contain much more important semantic information.
- The NE models that combine structure and content outperform the content-only and structure-only models ob-

Dataset	#tweet	#word	#hashtag	avgDocLen	avgHashtag
Tweet2011	333,491	12,420	106,682	5.22	1.42
Tweet2015	250,306	8,300	66,384	7.22	1.76

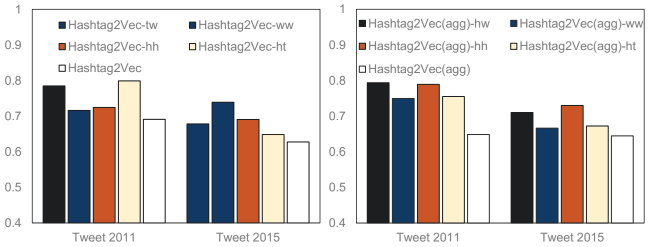
Table 1: Statisics of the two tweet collections.

type	Methods	Tweet2015	Tweet2011
Content only	NMF	0.7069	0.7166
	TWTM	0.715	0.9073
	TWDA	0.8876	0.938
	ATM	0.8612	0.7151
	HGTM	0.6909	0.6776
Structure only	LINE	0.9169	0.8554
	DeepWalk	0.9169	0.8503
	node2vec	0.9145	0.8639
Structure & Content	TADW	0.886	0.8664
	CANE	0.7665	0.7349
	Hashtag2Vec	**0.6276**	0.6915
	Hashtag2Vec(agg)	0.6447	**0.6488**

Table 2: Hashtag clustering performance on Tweet2011 & Tweet2015 datasets.

Methods	Tweet2015	Tweet2011
NMF	0.7575	0.8624
TWTM	0.2334	0.1344
TWDA	0.3787	0.2587
ATM	0.4533	0.594
HGTM	0.4493	0.5756
Hashtag2Vec	**0.1478**	0.1334
Hashtag2Vec(agg)	0.1494	**0.0934**

Table 3: Tweet clustering performance on Tweet2015 & Tweet2011 datasets.

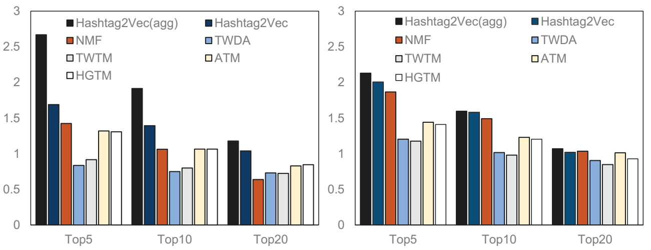


(a) Hashtag2Vec (b) Hashtag2Vec(agg)

Figure 2: Analysis of the Hastag2vec model components on Tweet 2011 and Tweet 2015.

viously, which confirms that the content and structure information is complementary.

- Our proposed models, Hashtag2Vec and Hashtag2Vec(agg) achieve the best performance on each dataset and boost the performance by a large margin.



(a) Tweet 2011 (b) Tweet 2015

Figure 3: Topic coherence comparison on Tweet 2011 and Tweet 2015.

4.5 Effectiveness of Model Components

We study the effectiveness of each component in Hashtag2Vec from the perspective of hashtag clustering. For evaluation, we train four degraded models of Hashtag2Vec: Hashtag2Vec-*tw*, Hashtag2Vec-*ww*, Hashtag2Vec-*hh*, and Hashtag2Vec-*ht*, each of which is obtained by removing the corresponding component defined in Eq. (3, 4, 9, 12) from the joint loss function, respectively. We compare the clustering performance of Hashtag2Vec with the four degraded methods to have an insight into the effectiveness of each component. Also, we conduct similar experiments on Hashtag2Vec(agg).

The performance comparisons are illustrated in Fig. 2. From the figures, the following observations can be obtained. (1) Removing each component results in a significant loss of performance on each dataset. This means that all the relations considered in our proposed Hashtag2Vec and Hashtag2Vec(agg) are semantically effective for embedding vector learning; (2) The tweet-word relation does not always contribute much for hashtag clustering, while the hashtag-word relation demonstrates more importance, which is because of the tweet aggregation alleviates the shortness and noisiness of the tweets; (3) The loss of performance caused by removing word-word relation is obvious in many cases. This phenomenon suggests that the word level embedding is an effective complement to the tweet and hashtag level content modeling. (4) Furthermore, it is learned that the hashtag co-occurrence plays an important role as shown by the drop of performances when it is removed from the model. Therefore, the random walk on hashtag co-occurrence network is proved to be effective in capturing their structural similarity. (5) Finally, the local structure also plays an important role in embedding vector learning, as the removal of hashtag-tweet relation results in significant performance degradation. This proves that there exist semantic similarities between hashtags and tweets in which they occur. And this kind of information can be learned via the local structure in Eq. (12).

In summary, the heterogeneous relations considered in Hashtag2Vec introduce effective semantic information. By combining the relations seamlessly via joint learning framework, Hashtag2Vec gains obvious improvements from each component.

(a) Results on Tweet2011

		SUPERBOWL	EGYPT	JOB	SONG
Tweet2011	WORDS	game bowl super steeler packer win commercial green fan team	mubarak egyptian egypt people protester protest cairo square police army	job manager sale engineer service hire senior developer project assistant	song listen feat album play black girl radio app ipod
	HASHTAGS	#superbowl #steelers #packers #nfl #sb45 #lfc #bears #jets #nba #reach	#egypt #jan25 #mubarak #tahrir #cairo #fb #p2 #25jan #tunisia #tcot	#jobs #mobsterworld #job #hiring #freelance #in #careers #indeed #marketing #it	#nowplaying #np #music #fb #musicmonday #itunes #video #blogtalkvideo #pandora #rock

(b) Results on Tweet2015

		CHARLESTONSHOOTING	YTFF	FATHERSDAY	QUOTE
Tweet2015	WORDS	shooting church people amp hate gun family victim black prayer	pick please see pls want youtubers dream help make come	day father happy dad amp gift love family idea enter	life never make people love tim fargo thing give good
	HASHTAGS	#charlestonshooting #charleston #prayersforcharleston #blacklivesmatter #prayforcharleston #rip #ameshooting #gunsense #nra #2a	#closeupatytff #ytff #cornettoatytff #rexonaatytff #sunsilkatytff #charlestonshooting #ytffmanila #chumfmmmva #yttf #mmvas	#fathersday #happyfathersday #father #win #giveaway #dad #presents #spoildad #love #insideout	#quote #quotes #life #leadership #inspiration #quoteoftheday #motivation #success #social #godfirst

Table 4: An illustration of topics. Each topic is shown with the top-10 words and hashtags that have the highest probabilities conditioned on that topic.

4.6 Topic Coherence Evaluation

Embedding models are expected to learn coherent topics to facilitate the semantic understanding. Hence we evaluate the topic coherence of word embedding vectors learned by Hashtag2Vec. We adopt PMI-Score [Newman *et al.*, 2010] for this evaluation, as it broadly agrees with human-judgment. PMI-Score calculates the average semantic relevance between top words under each topic, higher PMI-score indicates better coherence. Dimensions of the embedding space are taken as topics. Given the k-th topic, the M most probable words $(w_1^k, \cdots, w_M^k)$ can be picked by their weights on the k-th embedding dimension, and PMI-Score is defined as:

$$\text{PMI-Score} = \frac{1}{K \cdot \binom{M}{2}} \sum_{1 \leq i < j \leq M} \text{PMI}(w_i^k, w_j^k), \qquad (16)$$

where $\binom{M}{2}$ is the combination number of the top words, and K is the number of topics.

We compare the proposed Hashtag2Vec with other state-of-the-art topic models. Figure 3 shows the results on the Tweet2011 and Tweet2015 collections, and the numbers of most probable words are $5, 10, 20$, respectively. It can be observed that Hashtag2Vec outperforms the comparison methods on both datasets, and the Hastag2Vec(agg) model with aggregation performs even better.

4.7 Case Study

In order to intuitively show the quality of learned embeddings, we illustrate some topics in Table 4. They are SUPERBOWL, EGYPT, JOB, SONG for Tweet2011, and CHARLESTONSHOOTING, YTFF, FATHERSDAY, QUOTE for Tweet2015. It can be seen that these topics' hashtags and words are topic coherent and closely relevant. For example, it is interesting to see that the hashtag '#in', whose meaning is 'post to Linkedin', is highly ranked for the topic of 'JOB'. And for the topic of CHARLESTONSHOOING, it is closely related to '#nra' which means "national rifle association". These examples show that our model does learn high-quality topics through the joint embedding framework.

5 Conclusion

In this paper, we introduced a hashtag embedding method called Hashtag2Vec which incorporates heterogeneous information from the hierarchical network derived from short social texts. The proposed Hashtag2Vec leverages two-level content information and two kinds of structures, which effectively models the complex heterogeneous relations. Through a joint embedding framework, the embeddings of multiple kinds of nodes are learned simultaneously and enhanced mutually. Extensive experiments conducted on two real-world datasets from Twitter demonstrated the effectiveness and superiority of Hashtag2Vec.

Acknowledgements

This research is supported by the National Natural Science Foundation of China under the grant No. U1633103, the Key Projects in Tianjin Science and Technology Pillar Program Under the grant No. 17YFZCGX00610, and the Open Project Foundation of Information Technology Research Base of Civil Aviation Administration of China under the grant No. CAAC-ITRB-201601.

References

[Adamopoulos and Todri, 2015] Panagiotis Adamopoulos and Vilma Todri. The effectiveness of marketing strategies in social media: Evidence from promotional events. In *Proceedings of KDD*, pages 1641–1650, 2015.

[Church and Hanks, 1989] Kenneth Ward Church and Patrick Hanks. Word association norms, mutual information and lexicography. In *Proceedings of ACL.*, pages 76–83, 1989.

[Efron, 2010] Miles Efron. Hashtag retrieval in a microblogging environment. In *Proceeding of SIGIR*, pages 787–788, 2010.

[Godin *et al.*, 2013] Fréderic Godin, Viktor Slavkovikj, Wesley De Neve, Benjamin Schrauwen, and Rik Van de Walle. Using topic models for twitter hashtag recommendation. In *Proceedings of WWW*, pages 593–596, 2013.

[Gong and Zhang, 2016] Yuyun Gong and Qi Zhang. Hashtag recommendation using attention-based convolutional neural network. In *Proceedings of IJCAI*, pages 2782–2788, 2016.

[Grover and Leskovec, 2016] Aditya Grover and Jure Leskovec. node2vec: Scalable feature learning for networks. In *Proceedings of KDD*, pages 855–864, 2016.

[Huang *et al.*, 2010] Jeff Huang, Katherine Thornton, and Efthimis N. Efthimiadis. Conversational tagging in twitter. In *Proceedings of HT*, pages 173–178, 2010.

[Lee and Seung, 2000] Daniel D. Lee and H. Sebastian Seung. Algorithms for non-negative matrix factorization. In *Proceedings of NIPS*, pages 556–562, 2000.

[Levy and Goldberg, 2014] Omer Levy and Yoav Goldberg. Neural word embedding as implicit matrix factorization. In *Proceedings of NIPS*, pages 2177–2185, 2014.

[Li *et al.*, 2013a] Shuangyin Li, Guan Huang, Ruiyang Tan, and Rong Pan. Tag-weighted dirichlet allocation. In *Proceedings of ICDM*, pages 438–447, 2013.

[Li *et al.*, 2013b] Shuangyin Li, Jiefei Li, and Rong Pan. Tag-weighted topic model for mining semi-structured documents. In *Proceedings of IJCAI*, pages 2855–2861, 2013.

[Ma *et al.*, 2014] Zongyang Ma, Aixin Sun, Quan Yuan, and Gao Cong. Tagging your tweets: A probabilistic modeling of hashtag annotation in twitter. In *Proceedings of CIKM*, pages 999–1008, 2014.

[Newman *et al.*, 2010] David Newman, Jey Han Lau, Karl Grieser, and Timothy Baldwin. Automatic evaluation of topic coherence. In *Proceedings of NAACL*, pages 100–108, 2010.

[Perozzi *et al.*, 2014] Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. Deepwalk: online learning of social representations. In *Proceedings of KDD*, pages 701–710, 2014.

[Rosen-Zvi *et al.*, 2010] Michal Rosen-Zvi, Chaitanya Chemudugunta, Thomas L. Griffiths, Padhraic Smyth, and Mark Steyvers. Learning author-topic models from text corpora. *ACM Trans. Inf. Syst.*, 28(1):4:1–4:38, 2010.

[Sedhai and Sun, 2015] Surendra Sedhai and Aixin Sun. Hspam14: A collection of 14 million tweets for hashtag-oriented spam research. In *Proceedings of SIGIR*, pages 223–232, 2015.

[Shi *et al.*, 2016] Bichen Shi, Georgiana Ifrim, and Neil J. Hurley. Learning-to-rank for real-time high-precision hashtag recommendation for streaming news. In *Proceedings of WWW*, pages 1191–1202, 2016.

[Tang *et al.*, 2015] Jian Tang, Meng Qu, Mingzhe Wang, Ming Zhang, Jun Yan, and Qiaozhu Mei. LINE: large-scale information network embedding. In *Proceedings of WWW*, pages 1067–1077, 2015.

[Tu *et al.*, 2016] Cunchao Tu, Weicheng Zhang, Zhiyuan Liu, and Maosong Sun. Max-margin deepwalk: Discriminative learning of network representation. In *Proceedings of IJCAI*, pages 3889–3895, 2016.

[Tu *et al.*, 2017] Cunchao Tu, Han Liu, Zhiyuan Liu, and Maosong Sun. CANE: context-aware network embedding for relation modeling. In *Proceedings of ACL*, pages 1722–1731, 2017.

[Vicient and Moreno, 2015] Carlos Vicient and Antonio Moreno. Unsupervised topic discovery in micro-blogging networks. *Expert Syst. Appl.*, 42(17-18):6472–6485, 2015.

[Wang *et al.*, 2016] Yuan Wang, Jie Liu, Yalou Huang, and Xia Feng. Using hashtag graph-based topic model to connect semantically-related words without co-occurrence in microblogs. *IEEE Trans. Knowl. Data Eng.*, 28(7):1919–1933, 2016.

[Weston *et al.*, 2014] Jason Weston, Sumit Chopra, and Keith Adams. #tagspace: Semantic embeddings from hashtags. In *Proceedings of EMNLP*, pages 1822–1827, 2014.

[Yan *et al.*, 2013] Xiaohui Yan, Jiafeng Guo, Yanyan Lan, and Xueqi Cheng. A biterm topic model for short texts. In *Proceedings of WWW*, pages 1445–1456, 2013.

[Yang *et al.*, 2015] Cheng Yang, Zhiyuan Liu, Deli Zhao, Maosong Sun, and Edward Y. Chang. Network representation learning with rich text information. In *Proceedings of IJCAI*, pages 2111–2117, 2015.