

Hypergraph Label Propagation Network

Yubo Zhang,¹ Nan Wang,¹ Yufeng Chen,¹ Changqing Zou,² Hai Wan,¹ Xinbin Zhao,¹ Yue Gao^{1*}

¹BNRist, KLISS, School of Software, Tsinghua University, China ²Huawei Noah's Ark Lab
 {zhangyb17, cyf19}@mails.tsinghua.edu.cn, wangnanstc@yeah.net, {aaronzou1125, wan.whyhigh}@gmail.com,
 {zxb, gaoyue}@tsinghua.edu.cn

Abstract

In recent years, with the explosion of information on the Internet, there has been a large amount of data produced, and analyzing these data is useful and has been widely employed in real world applications. Since data labeling is costly, lots of research has focused on how to efficiently label data through semi-supervised learning. Among the methods, graph and hypergraph based label propagation algorithms have been a widely used method. However, traditional hypergraph learning methods may suffer from their high computational cost. In this paper, we propose a Hypergraph Label Propagation Network (HLPN) which combines hypergraph-based label propagation and deep neural networks in order to optimize the feature embedding for optimal hypergraph learning through an end-to-end architecture. The proposed method is more effective and also efficient for data labeling compared with traditional hypergraph learning methods. We verify the effectiveness of our proposed HLPN method on a real-world microblog dataset gathered from Sina Weibo. Experiments demonstrate that the proposed method can significantly outperform the state-of-the-art methods and alternative approaches.

Introduction

Deep learning methods have benefited from large-scale labeled datasets and have shown great superiority in many application scenario such as visual object detection (He et al. 2016), natural language understanding (Wang et al. 2016), *etc.* However, in many real-world situations, labeled data is much more costly than unlabeled data. Therefore, semi-supervised learning which models the classifier with the unlabeled data with the clustering or manifold assumption (Zhou et al. 2004) has attracted much attention in the past decade to solve this problem.

Label propagation (Zhu and Ghahramani 2002) is a classical and effective semi-supervised learning procedure. It propagates labels from the labeled data points to the unlabeled ones through various algorithms. Zhu *et al* (Zhu and Ghahramani 2002) proposed to propagate the labels along

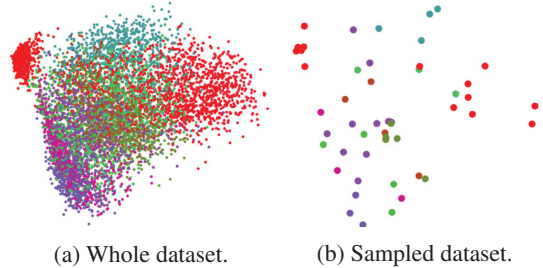


Figure 1: Distribution of the real-world hand-written digit dataset USPS. Left: distribution of the whole dataset; right: distribution of the sampled dataset.

the density of the whole data, including the labeled and unlabeled data, iteratively. Recently much research has been conducted on the label propagation algorithm, e.g., deploying short-cutting label propagation on the distributed system (Stergiou, Rughwani, and Tsioutsoulis 2018), adaptive label propagation through embedding (Zhang et al. 2018), applying to the fast-move data stream with a few labeled examples (Wagner et al. 2018), *etc.*

Label propagation algorithms roughly consist of connection construction of the graph and label propagation by random walk according to the propagation probability. There are two major challenges in these algorithms in real practice. The first challenge is how to construct effective connections among the data points. In real world, the distribution of either structural or non-structural data in the topological space may not purely satisfy the assumptions of manifold or clustering. Therefore it is essential to explore the optimal topological relevance structure where the two vertices connected in the graph likely belong to the same category. The second challenge is how to compute the propagation probabilities for the random walk in the graph. Although acceptable performances can be achieved in many cases with the uniform distribution assumption, optimizing the computation of propagation probability is challenging and can effectively improve the performance.

To address these two challenges in existing label propagation methods, we first introduce our intuitive thought. Fig.

*Corresponding author

1 shows the distribution of a real world handwritten digits dataset USPS (Hull 1994). As shown in Fig. 1a, many overlaps occur in samples belonging to some categories, and the distribution does not fit in a smooth manifold and clustering clearly. While the samples in Fig. 1b, which are a subsampled set of the whole dataset, may be shown easier to discriminate than the formers intuitively. Meanwhile, neural networks can automatically embed the samples from the origin space to a better metric space. Based on these hypotheses, we propose a Hypergraph Label Propagation Network (HLPN) method to conduct semi-supervised learning by considering the optimal structure. The whole scheme is illustrated in Fig. 2. Here we take the real-world multi-modality social network sentiment data in our experiments as an example, which includes three modalities, *i.e.*, text, vision and emoticon, for illustration. In this framework, each input batch consists of both labeled and unlabeled data. We first embed the input data to a feature space through a deep neural network. Then, we adopt a hypergraph module to explore the correlation information among the input samples. With the transductive label propagation from the labeled data to unlabeled ones on this hypergraph, we can obtain the predicted labels for all the unlabeled data in the batch. After that the cross-entropy loss can be computed and used to optimize the parameters in the embedding network. Through this way, the high-order relationship of data can be formulated in the hypergraph structure and the label propagation on the hypergraph can be conducted to predict labels for the unlabeled data.

The contributions of this paper are as follows:

- We propose a new hypergraph-based semi-supervised learning framework. This framework captures the manifold structure with sparse samples of the whole dataset and can achieve end-to-end learning. Our proposed method not only has the advantage of learning on the hypergraph structure but also is efficient, as it does not need to construct the hypergraph with all the samples of the whole dataset.
- We evaluate our method in a real-world multi-modality social network dataset crawled from the Sina Microblog Website. The results indicate that our method can improve the performance significantly compared with state-of-the-art methods.

The rest of this paper is organized as follows. We first introduce the related work. Then we present the detailed algorithm of HLPN. Experimental results and comparisons with state-of-the-art methods are provided in the following section. We finally conclude this paper.

Related Work

In many real-world applications, considering that the limited number of labeled samples may affect the performance of methods and that the information from abundant unlabeled data may improve the effectiveness, the semi-supervised method has attracted more and more attention from both industry and academic. Considering the superiority of exploring the local stationary structures, the research on graph-based methods has lasted for decades, and has been

widely used in many application, such as image classification (Dashtbozorg, Mendonca, and Campilho 2013), sentiment prediction (Ji et al. 2018), software defect prediction (Zhang, Jing, and Wang 2017), image retrieval (Xu et al. 2013) and so on.

Traditional graph-based methods set the unlabeled samples as the testing set and generate the labels of testing samples during the learning process, which can be divided into transductive learning methods and inductive learning methods. For the first type of methods, *i.e.*, transductive learning methods (Zhou et al. 2004; Chen et al. 2015), these methods need a retraining procedure for new test samples and cannot adapt to out-of-sample settings. Zhang *et al.* (Zhang, Jing, and Wang 2017) utilized a Laplacian score sampling strategy to construct a class-balance labeled training dataset and construct a nonnegative sparse graph. Then a label propagation algorithm is conducted to iteratively predict the labels of unlabeled samples. For the second type of methods, *i.e.*, inductive learning methods (Belkin, Niyogi, and Sindhvani 2006), these methods have the ability to classify the new testing samples. Jiang *et al.* (Jiang et al. 2017) defined a graph-based sparse prior and a sparse Bayesian model is obtained based on the traditional Bayesian inference technique, which is able to make decision in inductive ways.

However, it is known that the graphs structure which is constructed with features of labeled and unlabeled samples seriously affects the performance of the graph-based semi-supervised methods (Jebara, Wang, and Chang 2009). So it is meaningful to improve the effectiveness of graph structure and the features of samples in the learning process to avoid the impact of using low quality graph in graph-based semi-supervised methods. Jebara *et al.* (Jebara, Wang, and Chang 2009) proposed a b-matched graph to ensure each node in the graph has the same number of edges and the graph is exactly regular. Liu *et al.* (Liu et al. 2019) proposed a meta-learning method named transductive propagation network (TPN), which generated a graph structure to exploit the manifold structure in the data. In this method, the parameters of the feature embedding and the graph construction were optimized jointly. Li *et al.* (Li, Wang, and Zhou 2016) proposed a large margin separation method named LEAD to construct a safe graph-based method, which exploited large margin graphs while decreased the utilization of small margin graphs.

Although there are many works concentrating on graph-based semi-supervised methods, consider that in many real-world applications, the correlations among different samples are difficult to explore with the pairwise connection in graph structure. Hypergraph structure is proposed. The hypergraph structure is constructed according to the similarity between different samples, which is suitable for high-order relationship modeling. As for the construction of the hypergraph, it can be achieved by several ways. The most common way is based on the distances (Gao et al. 2012; Ji et al. 2018). In the construction process of hyperedges, each vertex is selected as centroid vertex each time. Then, the corresponding hyperedge is generated to connect the centroid vertex and the K nearest neighbors according to the distance. Hypergraph learning methods have been widely

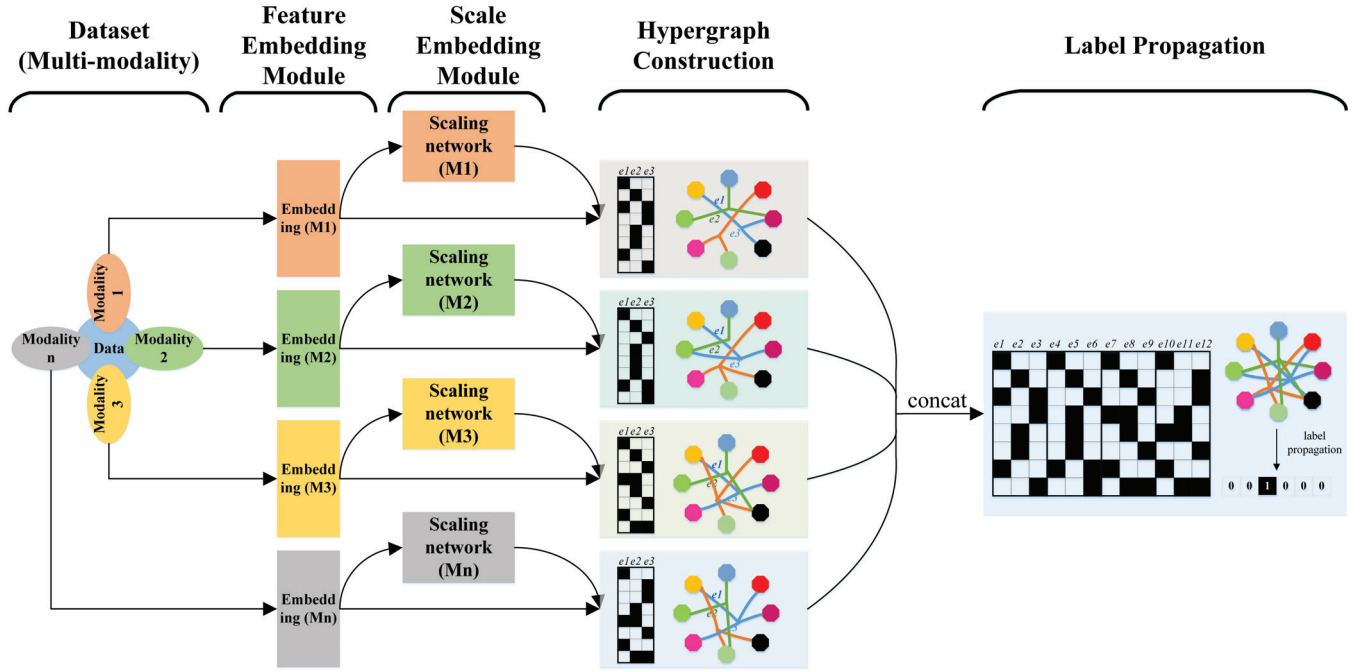


Figure 2: Framework of the proposed hypergraph label propagation network.

applied in many applications, such as image classification (Gao et al. 2012), segmentation (Ji et al. 2018) and so on. Ji *et al.* (Ji et al. 2018) proposed a bi-layer multi-modal hypergraph method for microblog sentiment prediction to avoid the influence of modalities missing. Moreover, in many real-world applications, obtaining labeled data is more difficult than gathering unlabeled data and the limited labeled samples may affect the effectiveness of the model. So it is important to use the unlabeled samples in the learning process. Thus, many semi-supervised methods (Chen et al. 2015; Gao et al. 2012; Feng et al. 2019; Jiang et al. 2019) based on hypergraphs have been proposed, which utilize the information of unlabeled samples by transductive learning. For example, Zhou *et al.* (Zhou, Huang, and Schölkopf 2007) employed the hypergraph structure to represent high-order relationships among the dataset, and extended the spectral hypergraph clustering method to hypergraph embedding and transductive classification. According to the above advantages, we constructed our method based on the hypergraph structure with both the parameters of the feature embedding and graph structure optimized simultaneously.

Hypergraph Label Propagation Network

In this section, we first introduce the typical hypergraph-based high-order relationship exploring method, and then present the details of the proposed hypergraph label propagation network.

High-Order Relationship Exploring via Hypergraph

Due to the superiority of the hypergraph on exploring high-order relationship among data, we first introduce the hypergraph structure to model the complex information among testing samples. Different from simple graph structures whose edges only connect two vertices, the hyperedges in a hypergraph can link more than two vertices leading to more flexible connections among the data and better complex relationship modelling performance.

Given a hypergraph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{W})$, it generally contains three components, *i.e.*, a hyperedges set \mathcal{E} , a vertex set \mathcal{V} and the weights of the hyperedges \mathbf{W} . We employ incidence matrix \mathbf{H} to represent the hypergraph structure, and the entry (i, e) of \mathbf{H} is defined as

$$h(i, e) = \begin{cases} 1 & \text{if } i \in e \\ 0 & \text{if } i \notin e \end{cases}, \quad (1)$$

which indicates the connection between the vertex i and hyperedge e .

Each vertex of the hypergraph is associated with a vertex degree $d(v)$, which is defined as $d(v) = \sum_{e \in \mathcal{E}} \omega(e) h(v, e)$, and for each hyperedge, its degree is defined as $\delta(e) = \sum_{v \in \mathcal{V}} h(v, e)$. Then, diagonal matrices \mathbf{D}_v and \mathbf{D}_e are used to denote the degrees of vertices and hyperedges.

Recent learning methods (Wagner et al. 2018) have been conducted on the hypergraph structure for different objectives, such as classification, embedding and so on. Zhou *et al.* (Zhou, Huang, and Schölkopf 2007) proposed a regularization framework for classification on the hypergraph

as $\arg \min_{\mathbf{F}} \{\lambda R_{emp}(\mathbf{F}) + \Omega(\mathbf{F})\}$. In the framework, λ is the trade-off parameter to balance the influence of the hypergraph structure regularizer $\Omega(\mathbf{F})$ and the empirical loss $R_{emp}(\mathbf{F})$. \mathbf{F} is the to-be-learned label matrix.

Generally, the empirical loss $R_{emp}(\mathbf{F})$ is defined as $R_{emp}(\mathbf{F}) = \|\mathbf{F} - \mathbf{Y}\|_{\mathbf{F}}^2$, where \mathbf{Y} is the label matrix of labeled samples. From the smoothness aspect, the regularizer $\Omega(\mathbf{F})$ on the hypergraph structure is defined as

$$\Omega(\mathbf{F}) = \frac{1}{2} \sum_{e \in \mathcal{E}} \sum_{u, v \in \mathcal{V}} \frac{w(e) \mathbf{H}(u, e) \mathbf{H}(v, e)}{\delta(e)} \left(\frac{\mathbf{F}(u, k)}{\sqrt{d(u)}} - \frac{\mathbf{F}(v, k)}{\sqrt{d(v)}} \right)^2, \quad (2)$$

where hypergraph laplacian Δ is defined as $\Delta = \mathbf{I} - \mathbf{\Theta} = \mathbf{I} - \mathbf{D}_v^{-\frac{1}{2}} \mathbf{H} \mathbf{W} \mathbf{D}_e^{-1} \mathbf{H}^T \mathbf{D}_v^{-\frac{1}{2}}$. Here, \mathbf{I} is the identity matrix.

Then, the regularization framework can be rewritten as

$$\arg \min_{\mathbf{F}} \left\{ \text{tr}(\mathbf{F}^T \Delta \mathbf{F}) + \lambda \|\mathbf{F} - \mathbf{Y}\|_{\mathbf{F}}^2 \right\}, \quad (3)$$

which can be solved directly by

$$\mathbf{F} = \left(\mathbf{I} + \frac{1}{\lambda} \left(\mathbf{I} - \mathbf{D}_v^{-\frac{1}{2}} \mathbf{H} \mathbf{W} \mathbf{D}_e^{-1} \mathbf{H}^T \mathbf{D}_v^{-\frac{1}{2}} \right) \right)^{-1} \mathbf{Y}. \quad (4)$$

Although the above traditional hypergraph learning methods have shown good performance in many tasks, they suffer from their high computational cost which limits their applications in practice.

Hypergraph Label Propagation Network

Here, we let $\{X_i, y_i\}, i = 1 \dots N$ denote the multimodal data, and $X_i = \{x_i^m\}, m = 1 \dots M$, in which N is the number of data in the whole dataset, M is the number of modalities and y_i is the label of X_i . The first l data are used for training while the next $u = N - l$ data are used for testing. One batch input data, i.e., N' samples, consist of two parts: N'_l labeled ones sampled from the training set and N'_u unlabeled ones sampled from the testing set.

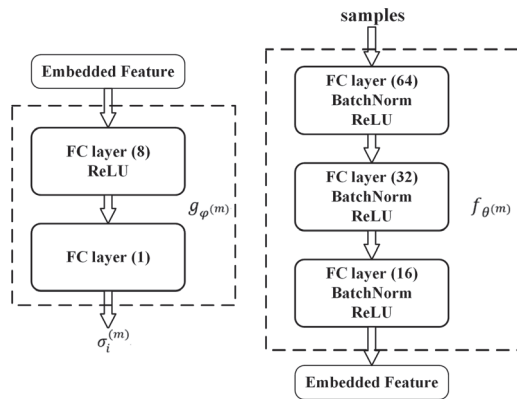


Figure 3: The deep neural network modules: Scale Embedding Module (left) and Feature Embedding Module (right).

The proposed network architecture is shown in Fig. 2. We first employ a feature extraction module which contains bag-of-words dictionaries of different modalities, the same as (Ji et al. 2018). By doing so, the output features are in 2547, 49 and 1200 dimensions with respect to the text, emoticon and visual modalities. We note that the feature extraction methods can be changed according to different tasks and the feature dimensions will be also changed then. In terms of the embedding module, we adopt multiple three-layer perceptions abbreviated as f_{θ^m} in which $f_{\theta^m}(x_i^m; \theta^m)$ maps the feature and θ^m is the parameter of the network for modality m . For simplicity, the number of neurons in the three layers of all modalities is set as 64, 32 and 16, as shown in the right of Fig. 3.

Inspired by (Liu et al. 2019), we propose an embedding module g_{ϕ^m} to inference the instance-wise scale parameter σ_i^m , as shown in the left of Fig. 3, which is used in computing the adjacency matrix:

$$\mathbf{A}^m = \exp \left(-\frac{1}{2} d \left(\frac{f_{\theta^m}(x_i^m)}{\sigma_i^m}, \frac{f_{\theta^m}(x_j^m)}{\sigma_j^m} \right) \right), \quad (5)$$

where $\mathbf{A}^m \in \mathbb{R}^{N' \times N'}$ is constructed among all the samples in one batch. This module not only can adjust the metric adaptively for each sample but also generate the proper probabilistic edge connection.

The hypergraph construction module aims to obtain predictions for the test set in batch, compute the cross-entropy loss and update the parameters in the modules before. In this module, we firstly compute the adjacency matrix \mathbf{A}^m for each modality, and construct a hypergraph based on the k-nearest neighbor method as

$$H^{(m)}(i, j) = \begin{cases} \mathbf{A}^m(i, j) & \text{if } i \in \mathbb{I}^m(j) \\ 0 & \text{if } i \notin \mathbb{I}^m(j) \end{cases}, \quad (6)$$

where $H^{(m)} \in \mathbb{R}^{N' \times N'}$, $\mathbb{I}^m(j)$ means the k-nearest samples of sample x_i^m in the metric space \mathbf{A}^m of modality m . In this way, we generate the hypergraph $\mathcal{G}^m = (\mathcal{V}^m, \mathcal{E}^m, \mathbf{W}^m)$ for the m -th modality. Note that we have conducted the optimization on the feature embedding module and the scale adjusting for each instance. To avoid over-engineering, we simply set the weights of hyperedges to be uniform. Then, we obtain M hypergraphs corresponding to the M modalities. These hypergraphs are next merged through simply concatenating to one large hypergraph, whose incidence matrix is $\mathbf{H} \in \mathbb{R}^{(M \times N') \times N'}$, to fuse the high-order topological information from different modalities. After that we conduct the transductive learning on the fused hypergraph, and according to Eq. 4, the closed-form solution \hat{F}_u for the predicted scores in one batch can be obtained. We then convert the score matrix \hat{F}_u to probabilistic score matrix through softmax

$$P(\hat{y}_i = j | X_i) = \frac{\exp(\hat{F}_{ij})}{\sum_{k=1}^{N'_u} \exp(\hat{F}_{ik})}, \quad (7)$$

where \hat{y}_i denotes the predicted label for the sample X_i and \hat{F}_{ij} is the predicting score for the sample X_i belonging to the j th category.

Algorithm 1 Hypergraph Label Propagation Network

INPUT: multi-modal data samples $\{\mathbf{X}_i\}_{i=1}^n$, class indicator matrix \mathbf{Y} associated with the labeled data. Hyperparameters including the depth and width of Multiple Layer Perception in the feature embedding module and scale embedding module, batchsize N' , and the number of nearest neighbors k in constructing the hypergraph in the hypergraph construction module.

OUTPUT: class matrix $\tilde{\mathbf{Y}}$ associated with testing samples in dataset.

TRAINING:

Step 1 Sampling: Uniformly sampling N' instances as one batch set B from the training set and split the N' instances into N'_l labeled set B_l and N'_u to-predict set B_u , which can be presented by $B = [B_l, B_u]$.

Step 2 Predicting: Input the batch set B and predict the labels for to-predict set B_u . All instances in B are sent to the first module, *i.e.*, feature embedding module $f_{\theta(m)}$, to obtain the embedded features $f_{\theta(m)}(x_i^{(m)})$, $i = 1 \dots N'$. m defines the modality of data. Then the features are input to scale embedding module $g_{\phi}^{(m)}$, which output a example-wise scale parameter $\sigma_i^{(m)} = g_{\phi}^{(m)}(f_{\theta(m)}(x_i^{(m)}))$. Then the hypergraph construction module generate the k-nearest neighbor hypergraph $H^{(m)}$ on the metric Eq. 5. After merging these hypergraphs from different modalities, conduct the transductive learning to predict the labels for the B_u .

Step 3 Updating: Computing the cross-entropy loss on ground-truth of B_u . Update the parameters in feature embedding module and scale embedding module.

TESTING:

Step 1 Sampling: According to the configuration on batchsize N' , uniformly sampling N'_l instances from the training set and N'_u from the testing set to form a batch.

Step 2 Inference: Predict the labels for the testing samples and repeat several times to cover all of the testing set.

END

We use the cross-entropy loss on the test set to update the paramters $\theta((m))$ and $\phi^{(m)}$ in feature embedding module $f_{\theta(m)}$ and scale embedding module g .

Experiments

The Dataset Description and Setting

To evaluate the performance of the proposed method, we have conducted experiments on a dataset from the Sina Weibo platform (www.weibo.com) (Ji et al. 2018). The dataset is collected from the daily top-10 hot topics in Weibo during Feb. 2014 to Apr. 2014, which consists of three components, *i.e.*, 71.7% tweets containing images, 99.8% blogs containing texts and 31.9% tweets containing emoticons. We utilize a ANP detector library named SentiBank to transform the low-level features of the Twitter image set into mid-level

features, and ANPs with confidence coefficients higher than 0.8 are employed.

After data cleansing and preprocessing (Ji et al. 2018), we got 5K tweets with clean labels in total, including 4196 positive samples and 1354 negative samples. For generalization, our experiments are conducted with 10 fold cross-validation, randomly selecting 4650 samples as a training set, 400 samples as a validation set and 500 sample as a testing set, and we compare the average performance on 10-fold the results of state-of-the-art methods for fair evaluation.

In detail, the experimental settings are configured as follows. In consideration that the size of the testing set is 500, and in order to keep consistency between the training and testing procedures, we set the size of to-predict set B_u in one batch as 500, same as that of the testing set. For the scale of labeled set B_l in one batch, we just set the size to 1500. For the testing procedure, the size of B_l is set to 1500, which is same as the training procedure.



Figure 4: Representative tweet examples of the microblog dataset from Sina Weibo. Each sample contains an image, expressions in Chinese, and emoticons.

Compared Methods

To evaluate the effectiveness of the proposed methods, we compare HLPN with several state-of-the-art methods, including Cross-media Bag-of-words Model (Wang et al. 2014), Multi-kernel SVM (Zhang et al. 2011), MHG method (Chen et al. 2015) and Bi-layer Multi-modal Hypergraph learning (Ji et al. 2018).

Experimental Results

Experimental results of all compared methods on the real-world microblog dataset crawled from Sina Weibo are shown in Fig. 5 and Table 1.

Performance Comparison with Different Modalities

The comparison results on different combinations of modalities are demonstrated in Table 1. According to the experimental results, HLPN achieves the best performance compared with other compared methods on multi-modal prediction. For instance, on the textual modality, compared with

Table 1: Performance comparisons with the state-of-the-art methods on multi-modal sentiment prediction. “T”, “V” and “E” represent textual, visual and emoticon modalities, respectively. “TVE” represents the combination of textual, emoticon and visual modalities. MK-SVM is based on multiple modalities which is unable to be conducted on a single modality.

Methods	Modality Combination						
	T	E	V	TE	TV	EV	TVE
CBM-NB(Wang et al. 2014)	0.592	0.902	0.593	0.716	0.582	0.722	0.716
CBM-LR(Wang et al. 2014)	0.658	0.813	0.533	0.858	0.600	0.756	0.799
CBM-SVM(Wang et al. 2014)	0.660	0.893	0.603	0.893	0.618	0.814	0.816
MK-SVM (Zhang et al. 2011)	-	-	-	0.859	0.521	0.859	0.862
MHG (Chen et al. 2015)	0.606	0.863	0.579	0.873	0.586	0.864	0.886
Bi-MHG (Ji et al. 2018)	0.683	0.903	0.611	0.895	0.695	0.899	0.900
HLPN	0.883	0.889	0.907	0.896	0.946	0.931	0.947

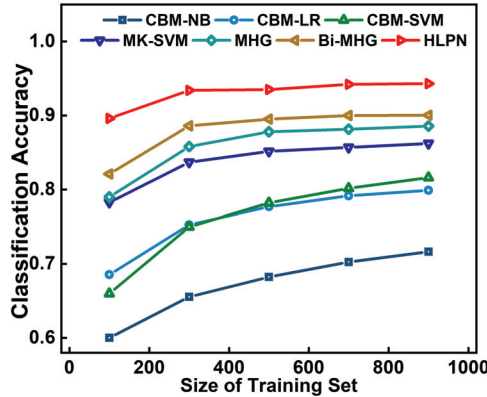


Figure 5: Performance comparison on different training data scales.

CBM-NB, CBM-LR, CBM-SVM, MK-SVM, MHG and Bi-MHG, HLPN achieves gains of 49.2%, 34.2%, 33.8%, 45.7% and 29.3%. On the visual modality, the proposed method achieves gains of 53.0%, 70.2%, 50.4%, 56.6% and 48.4%. On the emoticon modality, our method also shows better performance, which is comparable to Bi-MHG. The similar results can also be observed on the combination modalities. For example, on the combination of textual and visual modalities, HLPN achieves gains of 62.5%, 57.7%, 53.1%, 81.6%, 61.4% and 36.1% compared with the state-of-the-art methods. On the combination of textual, visual and emoticon modalities, the proposed method achieves gains of 32.3%, 18.5%, 16.1%, 9.9%, and 6.9% and 5.2%. These results justify that the HLPN is effective in exploring the relevance among the multi-modality dataset. Similar comparison results can be found on other combination of modalities.

Moreover, based on the comparison results, the performance of prediction on multi-modal schemes outperform the method on single modality schemes. For instance, for the combination of textual, visual and emoticon modalities, the proposed method achieves gains of 5.6% compared with the combination of textual and emoticon modalities and 1.7% compared with the combination of emoticon and visual

Table 2: Time-cost comparison with MHG and Bi-MHG on the combination of all three modalities.

Methods	Running Time
MHG (Chen et al. 2015)	178 seconds
Bi-MHG (Ji et al. 2018)	24 hours+
HLPN	36 seconds

modalities. Meanwhile, the performance from the visual and emoticon modality have been significantly improved in our method, and it is even higher than that from the emoticon modality, which indicates that our method can better understand the complex modality data than existing methods. The performance of the single textual modality is a little lower than that of the single emoticon modality, while the performance of the fused textual and visual modalities is higher than that of the emoticon and visual modalities, which indicates textual and visual modalities have more complementary information than emoticon and visual modalities.

Performance Comparison with Different Sizes of Training Data In order to evaluate the influence of different size of training data, we have conducted the comparison experiments with different sizes of training data, *i.e.*, 100, 300, 500, 700, 900, and the experimental results are shown in Fig. 5, which indicate the superiority of our method. For example, our method achieves gains of 49.3%, 29.8%, 27.0%, 25.4% and 24.0% compared with CBM-NB, and 9.1%, 5.4%, 4.5%, 4.7% and 4.7% compared with Bi-MHG on 100, 300, 500, 700 and 900 size of training sets. According to the comparison results, we also find that even with limited training data, our method achieves better performance. For instance, our method achieves 49.3%, 42.5%, 37.0%, 34.1% and 31.6% with the 100 sample training set, and 30.7%, 24.1%, 20.4%, 20.9% and 18.0% with the 200 sample training set. This result shows that the robustness of our method with scarce training samples.

In addition to evaluation on the accuracy, we also test the efficiency performance of our method. We count the testing time-cost of HLPN, compared with two effective methods, *i.e.*, Bi-MHG (Ji et al. 2018) and MHG (Chen et al. 2015). The results are illustrated in Table 2. It can be seen that our method is almost five times less time-consuming than MHG,

and compared with Bi-MHG, it reduces the time from more than one day to several seconds. These counts show that our method can also perform efficiently under the condition of superior effectiveness.

Ablation Study

In our model, hypergraph construction and the scale embedding module are the two main points, and our assumption is that the scale embedding module can improve the capacity of the hypergraph. Aiming to prove its effectiveness, we make an ablation study on the scale embedding module, *i.e.*, optimizing the scale σ_i^m through the scale embedding module *vs* setting the $\sigma_i^{(m)}$ as a hyperparameter. The results are illustrated in Fig. 6. For fair comparison, we tune the σ and then set it as 10.

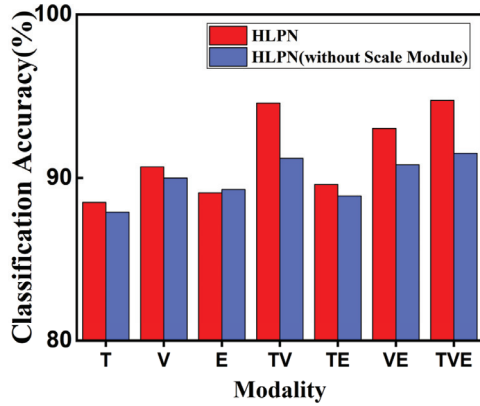


Figure 6: Performance comparison between HLPN and HLPN with the scale embedding module removed for each modalities.

From the results shown in Figure 6, we find that the Scale Embedding Module is helpful to the performance of our proposed method in most cases. Although it works not quite as obviously when we conduct it on only one modality, and especially reduces the performance only slightly in the Emoticon modality, it shows significant improvement when adopted on multiple modalities. With the combination of different modalities, such as Text and Visual modalities or others, the classification accuracy can improve about 3% or greater. It proves that the Scale Embedding Module is really effective in our method. From intuition, a subject can be represented by multi-modality, and the representation from different modalities is complementary. In our method, we suppose that the complementary information is latent in the topological relationship of some space, thus we adopt the hypergraph structure to mine the relationship. The Scale Embedding Module plays a key role in this work.

Discussion

It is illustrated in Table 1 that our proposed method achieves much better results in classification compared with other state-of-the-art methods and classical methods. As we can see, the results on the data from “Text” and “Visual” modalities improve significantly, and the result on the combination

of these two modalities also shows the effectiveness of our proposed method, while for the modality “Emoticon”, our method is not superior to other compared methods. This is due to the fact that feature extracted from the “Emoticon” modality are binary, *i.e.*, 0 or 1, and quite sparse as well. This property of the feature from “Emoticon” is limited in its capacity of representation. Under this situation, we can find that many methods can achieve comparable performance in classification and infer that the capacity of this feature may be the bottle-neck for the performance. Even so, the data of the “Emoticon” modality does not harm the results of “TVE” modality. This illustrates that the performance of our proposed method will not be reduced seriously when adding the data of another modality that is not of good quality.

The experimental results show that our method has a much better performance compared with the state-of-the-art methods, especially on the condition that the labeled data is limited. We analyze that it is due to the advantage that our model has the ability to optimize the representation of the samples on the topological space. In this architecture, the neural network can be updated gradually by selecting a sub-fraction of the dataset rather than the whole data.

Based on the assumption that the distribution of data is a smooth manifold, the traditional graph or hypergraph-based semi-supervised learning can be explained solidly and proved effectively. However, this hypothesis limits the capacity of these methods, because data in the real world is complicated, and may not satisfy a smooth manifold of high-quality, *i.e.*, supposing the smoothness of the data distribution may cause the loss of topological information. In our method we adopt a sub-fraction of the whole data to generate a sub-hypergraph and update the parameters of a neural network with the sub-hypergraph, *i.e.*, the effective metric on samples, to make the model more adaptive to the manifold over this subset of the whole dataset. Our proposed method samples the sub-structure of the whole manifold to exploit the sparse topological structure, or “sharp” manifold and memorizes this information to neural networks. Actually exploiting the latent manifold and optimizing the metric on hypergraph-based label propagation through traditional methods is difficult and a tall to mathematical requirement, while it can be achieved easily through the combination of label propagation and a deep neural network. Satisfactory performance on the real-world multi-modality data proves its reasonability.

Conclusion

In this work, we propose a Hypergraph Label Propagation Network (HLPN) that combines the hypergraph-based label propagation and deep neural networks, towards constructing the optimal topological hypergraph and conducting transductive label propagation in an end-to-end manner. We obtain the state-of-the-art result on a real-world multi-modal microblog tweets dataset for sentiment prediction.

Our future work will focus on two directions. The first one is to design an end-to-end architecture which optimizes the parameters of both feature embedding and hypergraph construction simultaneously and replaces the simple bag-of-words feature extractor. The second one is that our method

shows great advantages in multi-modality fusion. We will organize the model as a building block for the modalities fusion task.

Acknowledgements

This work was supported by the National Natural Science Funds of China (U1701262).

References

- Belkin, M.; Niyogi, P.; and Sindhvani, V. 2006. Manifold Regularization: A Geometric Framework for Learning from Labeled and Unlabeled Examples. *Journal of Machine Learning Research* 7(Nov):2399–2434.
- Chen, F.; Gao, Y.; Cao, D.; and Ji, R. 2015. Multimodal Hypergraph Learning for Microblog Sentiment Prediction. In *IEEE International Conference on Multimedia and Expo*, 1–6. IEEE.
- Dashtbozorg, B.; Mendoneca, A. M.; and Campilho, A. 2013. An Automatic Graph-based Approach for Artery/vein Classification in Retinal Images. *IEEE Transactions on Image Processing* 23(3):1073–1083.
- Feng, Y.; You, H.; Zhang, Z.; Ji, R.; and Gao, Y. 2019. Hypergraph Neural Networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, 3558–3565.
- Gao, Y.; Wang, M.; Tao, D.; Ji, R.; and Dai, Q. 2012. 3-D Object Retrieval and Recognition with Hypergraph Analysis. *IEEE Transactions on Image Processing* 21(9):4290–4303.
- He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep Residual Learning for Image Recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 770–778.
- Hull, J. J. 1994. A Database for Handwritten Text Recognition Research. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 16(5):550–554.
- Jebara, T.; Wang, J.; and Chang, S.-F. 2009. Graph Construction and b-matching for Semi-Supervised Learning. In *International Conference on Machine Learning*, 441–448. ACM.
- Ji, R.; Chen, F.; Cao, L.; and Gao, Y. 2018. Cross-modality Microblog Sentiment Prediction via Bi-layer Multimodal Hypergraph Learning. *IEEE Transactions on Multimedia* 21(4):1062–1075.
- Jiang, B.; Chen, H.; Yuan, B.; and Yao, X. 2017. Scalable Graph-based Semi-Supervised Learning through Sparse Bayesian Model. *IEEE Transactions on Knowledge and Data Engineering* 29(12):2758–2771.
- Jiang, J.; Wei, Y.; Feng, Y.; Cao, J.; and Gao, Y. 2019. Dynamic Hypergraph Neural Networks. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence*, 2635–2641.
- Li, Y.; Wang, S.; and Zhou, Z. 2016. Graph Quality Judgment: A Large Margin Expedition. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, 1725–1731.
- Liu, Y.; Lee, J.; Park, M.; Kim, S.; Yang, E.; Hwang, S. J.; and Yang, Y. 2019. Learning to Propagate Labels: Transductive Propagation Network for Few-shot Learning. In *International Conference on Learning Representations*.
- Stergiou, S.; Rughwani, D.; and Tsioutsoulouklis, K. 2018. Shortcutting Label Propagation for Distributed Connected Components. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*, 540–546. ACM.
- Wagner, T.; Guha, S.; Kasiviswanathan, S.; and Mishra, N. 2018. Semi-Supervised Learning on Data Streams via Temporal Label Propagation. In *International Conference on Machine Learning*, 5095–5104.
- Wang, M.; Cao, D.; Li, L.; Li, S.; and Ji, R. 2014. Microblog Sentiment Analysis Based on Cross-media Bag-of-words Model. In *Proceedings of International Conference on Internet Multimedia Computing and Service*, 76–80. ACM.
- Wang, Y.; Huang, M.; Zhao, L.; et al. 2016. Attention-based LSTM for Aspect-level Sentiment Classification. In *Proceedings of the 2016 conference on empirical methods in natural language processing*, 606–615.
- Xu, B.; Bu, J.; Chen, C.; Wang, C.; Cai, D.; and He, X. 2013. EMR: A Scalable Graph-based Ranking Model for Content-based Image Retrieval. *IEEE Transactions on Knowledge and Data Engineering* 27(1):102–114.
- Zhang, D.; Wang, Y.; Zhou, L.; Yuan, H.; Shen, D.; Initiative, A. D. N.; et al. 2011. Multimodal Classification of Alzheimer’s Disease and Mild Cognitive Impairment. *Neuroimage* 55(3):856–867.
- Zhang, Z.; Li, F.; Jia, L.; Qin, J.; Zhang, L.; and Yan, S. 2018. Robust Adaptive Embedded Label Propagation with Weight Learning for Inductive Classification. *IEEE transactions on neural networks and learning systems* 29(8):3388–3403.
- Zhang, Z.; Jing, X.; and Wang, T. 2017. Label Propagation Based Semi-Supervised Learning for Software Defect Prediction. *Automated Software Engineering* 24(1):47–69.
- Zhou, D.; Bousquet, O.; Lal, T. N.; Weston, J.; and Schölkopf, B. 2004. Learning with Local and Global Consistency. In *Advances in neural information processing systems*, 321–328.
- Zhou, D.; Huang, J.; and Schölkopf, B. 2007. Learning with Hypergraphs: Clustering, Classification, and Embedding. In *Advances in Neural Information Processing Systems*, 1601–1608.
- Zhu, X., and Ghahramani, Z. 2002. Learning from Labeled and Unlabeled Data with Label Propagation.