Event Classification in Microblogs via Social Tracking

YUE GAO, Tsinghua University
HANWANG ZHANG, National University of Singapore
XIBIN ZHAO, Tsinghua University
SHUICHENG YAN, National University of Singapore

Social media websites have become important information sharing platforms. The rapid development of social media platforms has led to increasingly large-scale social media data, which has shown remarkable societal and marketing values. There are needs to extract important events in live social media streams. However, microblogs event classification is challenging due to two facts, i.e., the short/conversational nature and the incompatible meanings between the text and the corresponding image in social posts, and the rapidly evolving contents. In this article, we propose to conduct event classification via deep learning and social tracking. First, we introduce a Multi-modal Multi-instance Deep Network (M^2DN) for microblogs classification, which is able to handle the weakly labeled microblogs data oriented from the incompatible meanings inside microblogs. Besides predicting each microblogs as predefined events, we propose to employ social tracking to extract social-related auxiliary information to enrich the testing samples. We extract a set of candidate-relevant microblogs in a short time window by using social connections, such as related users and geographical locations. All these selected microblogs and the testing data are formulated in a Markov Random Field model. The inference on the Markov Random Field is conducted to update the classification results of the testing microblogs. This method is evaluated on the Brand-Social-Net dataset for classification of 20 events. Experimental results and comparison with the state of the arts show that the proposed method can achieve better performance for the event classification task.

Categories and Subject Descriptors: I.4.8 [**Image Processing and Computer Vision**]: Scene Analysis; I.3 [**Image Processing and Computer Vision**]: Miscellaneous

General Terms: Algorithms

Additional Key Words and Phrases: Event classification, multi-modal, multi-instance, social tracking, Markov Random Field (MRF)

ACM Reference Format:
Yue Gao, Hanwang Zhang, Xibin Zhao, and Shuicheng Yan. 2017. Event classification in microblogs via social tracking. ACM Trans. Intell. Syst. Technol. 8, 3, Article 35 (February 2017), 14 pages.
DOI: http://dx.doi.org/10.1145/2967502

This research was supported in part by NSFC Program (No. 61671267, No. 91218302, No. 61527812, No. U1201251), National Science and Technology Major Project (No. 2016ZX01038101), MIIT IT funds (Research and application of TCN key technologies) of China, and The National Key Technology R&D Program (No. 2015BAG14B01-02).
Authors' addresses: Y. Gao and X. Zhao (corresponding author), School of Software, Key Laboratory for Information System Security, Ministry of Education, Tsinghua National Laboratory for Information Science and Technology (TNList), Tsinghua University, East Main Builing, Beijing, China, 100086; H. Zhang and S. Yan, National University of Singapore, 13 Computing Drive, Singapore, 117417.

DOI: http://dx.doi.org/10.1145/2967502

1. INTRODUCTION

Social media websites have shown their advantages in real-time information sharing in recent years [Fan and Gordon 2014]. The amount of social posts, such as microblogs, tweets, and blogs, is enormous and still rapidly increasing, which makes social media platforms more important for information transmission [Kaplan and Haenlein 2010], and thus has attracted much research attention recently [Asur and Huberman 2010; Gao et al. 2016]. Generally, the users in social media platforms are active and the information can be transmitted widely in a very short period. For example, Sina Weibo[1] has 129.1 million active users for each month and more than 100 million microblogs each day. The users may talk about real-life events and the live information may spread quickly and widely across the entire social network. For instance, the Super Bowl in 2013 attracted nearly 24 million tweets in total and the tweets about the blackout was over 0.2 million each minute. Another example is the MH370 event recently. MH370 has attracted a lot of attention and has become a hot topic these days. It is important to detect these events for live social media stream [Hearst 2009; Sayyadi et al. 2009; Zhou and Chen 2014] so further investigation and analysis can be conducted.

The microblog event relates to the events which attract attention from the social media users. It is noted that event detection in social media platforms is a challenging task due to the following facts. First, social media posts tend to be short and conversational. For example, social posts are around 140 characters or less in Twitter and Sina Weibo. The short social posts limit the useful content for processing. A single microblogs may be too short to convey adequate content for analysis. Second, the social media data evolve very fast. The contents for one event may update rapidly and the hot vocabularies change fast. Third, with the increasing heterogeneous and multimedia social posts, the text and the corresponding visual content may not have compatible meanings. At the same time, it is difficult to precisely annotate whether the image or the text is relevant to an entity for large-scale microblogs, which leads to a multi-modal (the textual and visual content) multi-instance (the individual textual and visual classification) weakly-labeled scenario.

In this work, we propose a microblog event classification method with a Multi-modal Multi-instance Deep Network (M^2DN) and social tracking. Deep learning has shown its superior ability for data representation and has been widely used in multiple areas, such as computer vision [Karpathy et al. 2014] and audio analysis [Pandey and Lazebnik 2011]. Convolutional neural network (CNN) [Kavukcuoglu et al. 2010; Krizhevsky et al. 2012a] is a typical deep learning framework which has been employed in visual object recognition and image classification. To tackle the incompatible meaning problem between the text and the image in microblogs, M^2DN is a two-pathway deep network, in which the two pathways process the text and the visual content, respectively. M^2DN is designed to handle the weakly-labeled multi-modal microblogs. Concerning the short and conversational microblogs nature, we propose to employ social information to enhance the M^2DN classification results. The objective of social tracking is to find a set of social-related microblogs, which have strong relations to the testing microblogs and further correct the single microblogs classification results. This correction is achieved by formulating all selected microblogs and the testing samples in an Markov Random Field (MRF) model and conducting inference on MRF. This method has been evaluated in the Brand-Social-Net dataset on the classification of 20 events. Experimental results demonstrate the effectiveness of the proposed method.

The remainder of this article is organized as follows. Section 2 reviews related work. Section 3 introduces the proposed microblogs event classification method. Section 4

[1]http://www.weibo.com.

provides the experimental results, followed by conclusions and discussions for further work in Section 5.

2. RELATED WORK

In this section, we briefly review the related work in event detection in social media platforms. Given new coming data, the similarity between the new data and the existing events are computed first and the event with the maximal similarity is selected. When all the similarities are below a predefined threshold, it will be considered as a new event. A modified term frequency/inverse document frequency (TF/IDF) and time-based threshold are employed in Allan et al. [1998] to measure the relevance between events and documents, in which an auxiliary dataset is used to estimate the IDF, due to the fact that the future documents are unknown. An improvement is proposed in Yang et al. [1998], in which an incremental IDF is introduced which considers a time window and a decay factor to measure the similarity between documents and events. Kleinberg [2002] propose to detect events by using an infinite state automaton and the events are formulated with state transitions. Pui et al. [2005] introduce the exploration of word appearance as the binomial distribution, and the word burst is identified by a heuristic with thresholds. The frequency domain of text content has also been investigated. He et al. [2007] introduce the Discrete Fourier Transformation (DCT) to detect the burst in the time domain by using the spike in the frequency domain. The Wavelet-based signal processing has been introduced in Weng and Lee [2011] to detect events, in which the cross-correlation between the word appearance is measure by using the Wavelet-based feature.

Reuter and Cimiano [2012] propose an event classification method dealing with incremental data in social media streams. In this method, first a candidate retrieval step is performed to gather related events by using the capture time, upload time, geographic location, tags, titles, and the description. Then, the similarities between one document and one event for the top returned retrieved events are measured based on nine features, such as the temporal information, geographical information, and textual information. Then, the probabilities of the documents belonging to the event or belonging to a new event can be computed by a trained Support Vector Machine. A threshold is empirically selected using a gradient descent method on a split of training data. Becker et al. [2010] introduce learning similarity metrics to identify events in social media streams, in which the event identification task is formulated as a clustering problem. In this method, each event is denoted by a document cluster, and the scalable clustering is evaluated using normalized mutual information [Strehl et al. 2002; Manning et al. 2008] and B-cubed [Amigó et al. 2009]. Considering the different information in social media documents, such as the textual feature and the location data, different similarities combined in an ensemble-clustering procedure. To classify new data into the existing events, a group of training samples are firstly selected from labeled data, and the logistic regression and Support Vector Machine (SVM) are employed as the classifier, which shows the best performance in experimental results, i.e., CLASS-LR and CLASS-SVM.

As social posts are generally short, it is hard to directly employ keywords-based methods, and the traditional bag-of-words methods also have limitations. Sriram et al. [2010] proposed to employed a set of domain-specific features from the authors' profile and text content. The tweets can be classified into some general categories by using the Naive Bayes classifier. It is noted that most of the existing methods conduct classification on a single microblogs directly. Due to the two challenges as introduced in Section 1, it is necessary to further explore the microblog relations to enrich the short social posts and jointly learn the text and the visual contents.

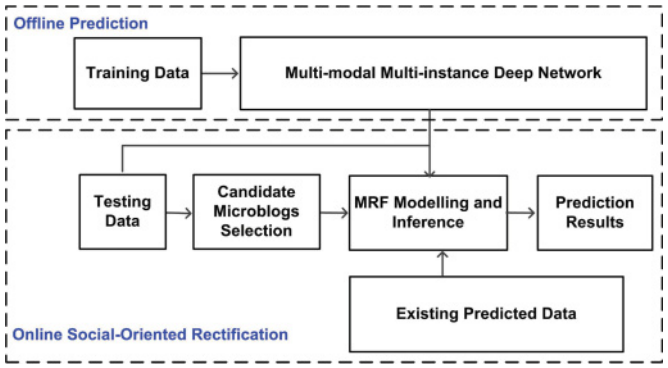

Fig. 1. The framework of the proposed event classification method in microblogs.

It is noted that most of the existing methods are based on the texture content associated with the timestamp. With the increasing multimedia content in social media streams, such as images and videos, it is important to further explore the visual context in a microblog for event analysis.

3. THE PROPOSED EVENT CLASSIFICATION METHOD

In this section, we introduce the proposed event classification algorithm. As described in Figure 1, the proposed method is composed of two parts. In the single microblogs classification part, an M^{2}DN is trained for the event, and the testing data can be first classified with the classifier. In the social tracking part, a set of social-related microblogs are selected first, and then they are formulated in an MRF model with the testing data. The inference on MRF determines the final event classification results.

3.1. Event Classification by M^{2}DN

Many microblogs Application Programming Interface (API) allow us to conveniently harvest microblogs according to a certain brand at a large scale, however, it is still prohibitively expensive to precisely label whether the text or the corresponding image is related to a brand in microblogs. Hence, we eventually result in positive or negative samples at the microblog-level, containing multiple instances of multi-modality, e.g., image and text. In nature, it can be formulated into a multiple instance classification problem [Maron and Ratan 1998; Babenko et al. 2009] since the training samples are only weakly labeled at the "bag-of-instances" level. It is worth noting that the instances in microblogs are multimodal, which should be first transformed into a homogeneous feature space, e.g., using canonical correlation analysis (CCA) [Thompson 2005; Hardoon et al. 2004]. Unfortunately, the images and text in microblogs are not necessarily correlated, and thus, it is generally intractable to learn a common space by traditional shallow methods like CCA. We proposed a novel deep learning architecture, namely, M^{2}DN, to jointly learn the common features of both modalities and infer the exact relatedness of the multimodal multiple instances. Next, we will introduce the architecture overview and the learning details.

3.1.1. Architecture Overview. As shown in Figure 2, M^{2}DN constitutes: (a) two pathways, each of which connects to one single input modality, i.e., image or text, (b) an aggregation layer, where each of the unit is only activated by the most responsive output units of the two pathways, and (c) a C-way softmax layer which produces a distribution over the C brand labels. The objective of M^{2}DN is to maximize the average across training cases of the log-probability of the correct label under the prediction distribution. In

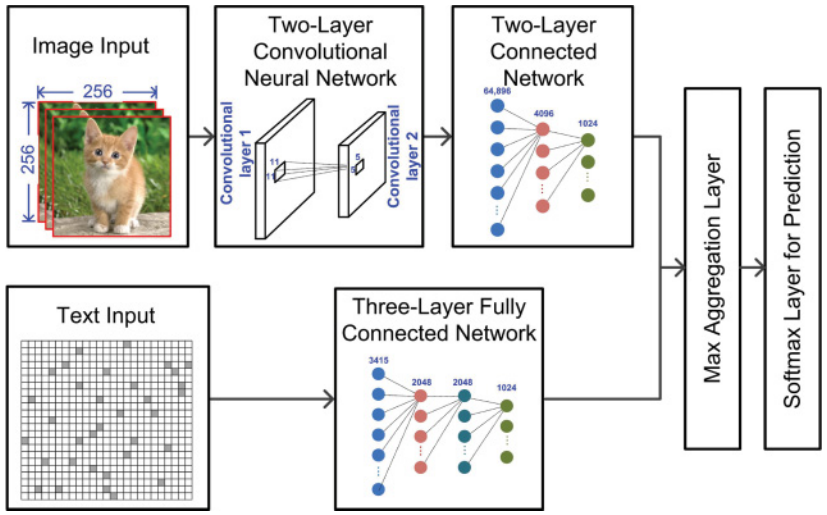

Fig. 2. The architecture of the M^{2}DN method.

particular, we expect that the softmax layer would be able to optimize the nominations of whether image or text contribute most to predict a brand from the aggregation layer.

The image pathway contains two convolutional layers (followed by max-pooling) and two fully connected layers. Since the visual patterns of brand images are structural, the locally shared weights of convolutional layers allow our network to learn different high-level features at different locations, facilitating better captures of brand structures. The first convolutional layer filters the $256 \times 256 \times 3$ color input images with 96 kernels of size $11 \times 11 \times 3$ with a stride of 4 pixels. The second convolutional layer takes as input the output of the first convolutional layer (with a size of 2×2-pixel and overlap of 1-pixel max-pooling), and filters it with 256 kernels of size $5 \times 5 \times 48$. This results in 64,896 neurons as the output of the second convolutional layer. The two fully connected layers transform the 64,896 inputs into subsequent 4,096 and 1,024 outputs, respectively. Note that the neuron activation functions of the two convolutional layers and the first fully connected layer are a rectified linear unit [Krizhevsky et al. 2012b] and that of the last fully connected layer is a sigmoid unit. The text pathway contains three fully connected layers, which start from the 3,415 textual input in terms of term presence of each tweet. Then, we transform the input into 2,048; 2,048; and 1,024 outputs by the subsequent three layers, respectively.

Yet, we do not know whether the image or the tweet is related to the target brand. Hence, we choose the maximum output neuron activations from both pathways to support the final prediction. Formally, for each neuron z_i of the aggregation layer, we have $z_i = max(x_i, t_i)$, where x_i and t_i are the ith neuron of the image and text pathway, respectively. By doing this, we expect the overall network can infer the most confident activation with respect to the target brand by feed-forward inference and learn modality-specific brand beliefs through back-forward training.

3.1.2. Pre-training. It is well known that deep architecture only works well if the trainable parameters are properly initialized to a good solution. In this section, we introduce how to use the Restricted Boltzman Machine (RBM) [Bengio 2009] to pre-train the proposed M^{2}DN.

Since the brand occurs in natural user-uploaded images in real applications, it is crucial for our network to recognize generic image patterns, such as edges and parts

for further learning the brand representations. Therefore, for the two convolutional layers of the image pathway, we use the trained model of the first two convolutional layers in Donahue et al. [2013], since the model specifications of these two layers are identical. The pre-trained convolutional layers encode valuable lower-level visual patterns acquired from large-scale ImageNet datasets, and hence, they are expected to help us in capturing generic visual patterns. For pre-training the two fully connected layers, we adopt an autoencoder [Bengio 2009] which can initialize the parameters in an unsupervised way. In particular, for each layer, we train the parameters by minimizing the reconstruction error of the decoded signals and the original input signals.

For pre-training the text pathway, we adopt the RBM [Bengio 2009] to layer-wisely initialize the parameters, since it is proved to be effective for sparse textual features. The RBM is an undirected graphical model that connects two layers of random variables. Without the loss of generality, we denote $\mathbf{v}$ as visible variables and $\mathbf{h}$ as hidden variables corresponding to any two connected layers in M^{2}DN. In particular, for the combined image and text variables at Layer 2, we have $\mathbf{v} \leftarrow [\mathbf{x}^{2^T}, \mathbf{t}^{2^T}]^T$, and for the shared hidden layer, we have $\mathbf{h} \leftarrow \mathbf{z}$. The connections in RBM are parameterized by $\mathbf{W}$ between $\mathbf{v}$ and $\mathbf{h}$. The optimization is to minimize the negative logarithm of the likelihood $p(\mathbf{v}) = \sum_{\mathbf{h}} p(\mathbf{v}, \mathbf{h}) = \sum_{\mathbf{h}} \frac{e^{-E(\mathbf{v},\mathbf{h})}}{Z}$, where Z is a partition function. By assuming $\mathbf{v}$ and $\mathbf{h}$ are $\{0, 1\}$-valued binary, conditionally independent variables, the energy function and its induced posterior probability are formulated as

$$\begin{cases} E(\mathbf{v}, \mathbf{h}) = -\mathbf{b}^T\mathbf{v} - \mathbf{c}^T\mathbf{h} - \mathbf{h}^T\mathbf{W}\mathbf{v}, \\ p(h_i = 1|\mathbf{v}) = \sigma(c_i + \mathbf{W}_{i\cdot}\mathbf{v}), \\ p(v_i = 1|\mathbf{h}) = \sigma(b_i + \mathbf{W}_{\cdot i}\mathbf{h}), \end{cases} \tag{1}$$

where $\mathbf{b}$ and $\mathbf{c}$ are model biases, $\mathbf{W}_{i\cdot}$ or $\mathbf{W}_{\cdot i}$ denotes the ith row or column of $\mathbf{W}$. Note, these probabilities are consistent with the sigmoidal activation functions of the three-layer textual pathway. As it is intractable to compute the gradient of the log-likelihood, we learn the parameters of the RBMs using contrastive divergence as in Tieleman [2008]. Moreover, in order to learn sparse models in an efficient way, we set the initial biases of the RBMs as sufficiently small (e.g., -2).

3.1.3. Fine-tuning. After pre-training M^{2}DN as above, we are able to fine-tune the entire network using stochastic gradient descent to minimize the objective function of M^{2}DN. As mentioned before, we want to maximize the average across training cases of the log-probability of the correct label under the prediction distribution by minimizing a C-way softmax function. Suppose the overall objective function (with ℓ_2-norm weight decay) is F, we update the model parameter $\mathcal{W}$ using the stochastic gradient descent, in which the update rule of $\mathcal{W}^i$ in the kth iteration is

$$\begin{cases} \Delta_{k+1} = 0.9 \cdot \Delta_k - 1.5e^{-4} \cdot \eta \cdot \mathcal{W}_k^i - \eta \cdot \frac{\partial F}{\partial \mathcal{W}_k^i}, \\ \mathcal{W}_{k+1}^i = \Delta_{k+1} + \mathcal{W}_k^i, \end{cases} \tag{2}$$

where Δ is the momentum variable [Qian 1999], and η is the learning rate which is adaptive to the objective function value.

With the trained M^{2}DN, we can predict a testing microblogs for the brands.

3.2. Social Tracking

Due to the short and conversational nature of microblogs and the rapidly evolving content, it is not precise to analyze each microblogs individually. For example, a classifier trained with the first week's data may not work well on the eighth week's testing data. Under such circumstance, it is difficult to identify related microblogs confronting little useful information and too much noise. Therefore, exploring other related resources

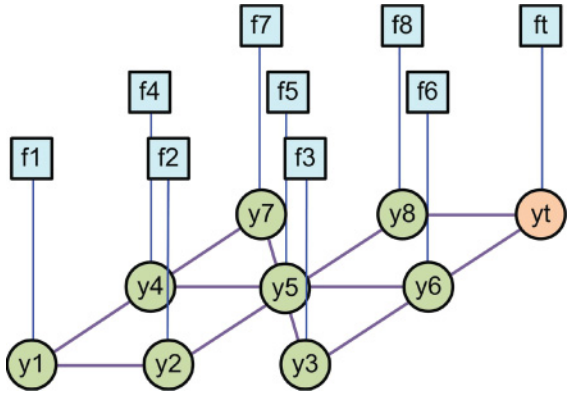

Fig. 3. Illustration of the MRF modeling.

regarding the target becomes an important way on this issue. We note that the microblogs in social media platforms are connected, and a possible solution to tackle this problem is to consider a social tracking for the testing data, which aims to find a set of microblogs linked to the testing data and further be used as the auxiliary information to update the testing data. The keywords for the event may vary with respect to different time, which makes it difficult to identify the later data to the event. In this case, a set of relevant microblogs in early time can help to keep tracking the testing data.

Here, we first briefly introduce the proposed social tracking method. When a new microblogs m_t comes, we first find the social-related microblogs M^t for m_t, which can be the comment information, repost information, and user connections in a time window. Then, an MRF is generated by using M_t and m_t together. In this MRF, the connection is constructed by using the textual/visual similarity between microblogs and the initial relevance scores for M^t are provided. For m_t, the initial relevance score comes from the M^2DN classifier. The inference on MRF further updates the relevance score $\mathbf{z}$ for m_t.

3.2.1. Social-Related Microblogs Selection. We first extract a microblog set M^t which can cover most of related microblogs with m_t with a constriction of time window, which is set as 2 weeks in this work. There are two conditions that the candidates should follow:

—The comments/repost information. It is possible to track the direct social connection of m_t, such as m_t is a comment or a reposted one for another microblogs. All these microblogs are the *parent* nodes for m_t, which have high possibility to be with similar content with m_t.
—User connection. Studies have shown people are often influenced by others and tend to follow the crowd [Easley and Kleinberg 2010]. On social media platforms, the user usually pays attention to what friends are talking about and will comment or repost the microblog from friends or post new microblogs on the similar topics. Hence, here we retrieve the microblog from the user's friends.

3.2.2. Relevance Inference on MRF. To reinforce the off-line M^2DN prediction results, the social path learning output M^t is employed as the auxiliary information here. In M^t, each microblogs is either with a rectified prediction score computed previously or from the ground truth. We use the relationship between microblogs and the prediction score of existing microblogs to infer and update the relevance score of m_t to all brands. Figure 3 illustrates the inference on the MRF.

With M^t, we construct an MRF to conduct the inference. Vertices correspond to the microblog in $M^t \cup m_t$. Here, the textual and visual content (if available) are taken into

consideration. For two microblogs m_i and m_j, the similarity in MRF is defined as:

$$\rho(m_i, m_j) = max\{\rho_v(m_i, m_j), \rho_t(m_i, m_j)\}, \tag{3}$$

where $\rho_v(\cdot)$ and $\rho_t(\cdot)$ are the visual and text similarities. In the case of the microblog having no text or images, ρ_t or ρ_v is set as 0. Here, we extract text information by employing the bag-of-words feature as mentioned earlier. $\rho_t(\cdot)$ is defined as the cosine similarity. For the visual information, we extract spatial pyramid feature [Lazebnik et al. 2006] and further reduce the dimension by Principle Component Analysis (PCA), denoted by $\mathbf{h}$. Then, the visual similarity between two microblogs is defined as

$$\rho(\mathbf{h}_i, \mathbf{h}_j) = \exp\left(-\frac{\|\mathbf{h}_i - \mathbf{h}_j\|_2^2}{2\sigma_v^2}\right), \tag{4}$$

where σ_v^2 is empirically chosen as the mean value of all image distances.

With the similarity defined by Equation (4), two vertices are connected if ρ is greater than a threshold.

After constructing the MRF, now we define the configuration of MRF. The latent variable $y = 1, 0$ stands for whether a microblog belongs to the event or not. The observation probability is defined as

$$p(f_i|y_i = 1) = \frac{1}{1 + e^{-\lambda z}}, \tag{5}$$

where f_i denotes the visual and text feature of the microblog, z denotes the classification score on the event, and λ is an empirically chosen constant. $p(f_i|y = 0) = 1 - p(f_i|y_i = 1)$. Equation (5) means the likelihood of the current microblogs is determined by the visual and text information and the M^{2}DN classification process. Here, the edge potential of two connected vertices is defined as

$$p(y_i, y_j) = \begin{cases} e^{-\beta} & \text{if } y_i = y_j \\ e^{\beta} & \text{if } y_i \neq y_j \end{cases}, \tag{6}$$

where β is a empirically chosen constant. This model means the two connected vertices tend to have consistent state.

Let $\mathbf{y} = (y_1, \ldots, y_N)$ and $\mathbf{f} = (f_1, \ldots, f_N)$, our objective is to optimize

$$p(\mathbf{y}|\mathbf{f}) = \prod_i p(f_i|y_i) \prod_{i,j} p(y_i, y_j) \tag{7}$$

We use the loop belief propagation method [Ihler et al. 2005] to do this optimization. After optimization, the returned belief for each brand is obtained.

4. EXPERIMENTAL RESULTS

In this part, we introduce the experiments, including the dataset and the experimental results.

4.1. The Dataset

To evaluate the proposed method, we conduct experiments in the Brand-Social-Net dataset [Gao et al. 2014], which is the first large-scale brand-related social media dataset. This dataset contains 20 saga events as listed in Table I. These saga events happened during June and July 2012, and the number of relevant microblogs for each saga event ranges from hundreds to thousands. Given the microblog of each saga event, event detection is conducted to explore the events in the saga event.

Table I. The 20 Saga Events in the Testing Dataset

Topic name	Short
The Apple Worldwide Developers Conference	DC
Windows 8 Preview	Win8
Office 2013 Preview	Office
Nokia Lumia	LU
Pepsi Michael Jordan Memorial Can	MJ
Samsung Galaxy 3 I9300	G3
HTC ONE	ONE
ShenGangMaco Auto Expo	SGM
Chongqing Auto Expo	CQ
Changchun Auto Expo	CC
Dior Addict	ADD
Hyundai MD Avante	AVA
Citroen DS5	DS5
Farrier Berlinetta F12	F12
Chrysler 300C 2012	300C
Honda Elysion	ELY
Honda CR-Z	CRZ
Mazda CX-5	CX5
Audi Q3	Q3
Toyota Highlander 2012	HI

4.2. Compared Methods

To evaluate the performance of the proposed method, the following methods are selected for comparison.

—The Candidate-Ranking (CR) method [Reuter and Cimiano 2012]. The CR method first retrieves several promising events and the probability of the incoming document for these events or a new event can be measured by an SVM classifier.
—The CLASS-SVM (CS) method [Becker et al. 2010]. CS is an incremental clustering method which employs SVM as the classifier to identify whether a new document belongs to an existing event or a new event.
—SVM classifier with text and visual feature (SVM).
—SVM classifier with social tracking using MRF (SVM+ST). In this method, SVM is employed to replace the M^2DN method for event classification.
—The M^2DN method without social tracking using MRF.
—The M^2DN method with social tracking using MRF, that is, the proposed method (M^2DN+ST).

In our experiments, we take the first 200 microblogs for each event as the training data which are labeled for the 20 events and leave all other data as the testing samples, which are used in the temporal order to simulate the live social media streams. Each testing microblogs is processed to predict it for the 20 events. In our experiments, Recall, Precision, and F measure are employed as the evaluation criteria, which measures the data coverage, the prediction accuracy, and the joint performance of Recall and Precision, and are defined as follows.

—*Recall* (Re). Recall measures the data coverage of event classification, and it is calculated by:

$$Re = \frac{\#\ Correct\ Event\ Classification}{\#\ All\ Relevant\ Microblogs},$$

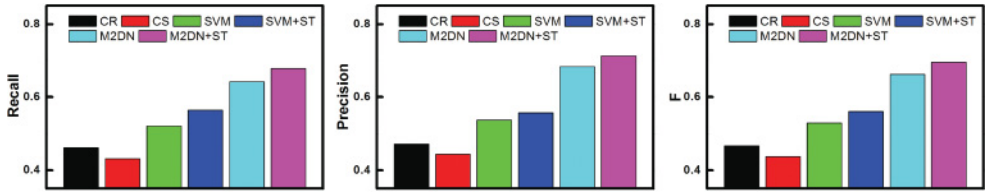

Fig. 4. The comparison of different methods on event classification in microblogs.

where # *Correct Event Classifications* is the number of corrected predicted microblogs for the event, and # *All Relevant Microblogs* is the number of relevant microblogs in the dataset for the event.

—*Precision* (Pr). Pr measures the accuracy of the event classification performance.

$$Pr = \frac{\#\ Correct\ Event\ Classifications}{\#\ All\ Classifications},$$

where #*All Classification* is the number of predictions for the events.

—*F-Measure* (F). F jointly considers *Recall* and *Precision*, which is defined as:

$$F = \frac{2 \times \mathrm{Re} \times \mathrm{Pr}}{\mathrm{Re} + \mathrm{Pr}}$$

4.3. Experimental Results and Comparisons

In this part, we provide the experimental results and further compare the proposed method with the state-of-the-art methods, i.e., CR [Reuter and Cimiano 2012], CS [Becker et al. 2010], and other baselines. Experimental results of different methods are provided on Figure 4 in terms of Recall, Precision, and F, respectively. As shown in these results, the M^{2}DN+ST method achieves the best results in comparison with all other methods. Here, we take the F measure at the 8th week as an example. The M^{2}DN+ST method achieves an improvement of 17.2%, 16.1%, 41.5%, 24.5%, and 6.5% compared with CR, CS, SVM, SVM+ST, M^{2}DN, respectively. Similar results can be found in the comparison of other time and evaluation measures. These results demonstrate that the proposed method is effective on brand prediction in microblogs.

We also show the results of four events in Figure 5. In all events, we can observe that the proposed method is able to achieve a large improvement on recall while maintaining a good precision. This improvement can be dedicated to two reasons. First, M^{2}DN is able to learn a better classifier which can handle the multi-modal multi-instance weakly labeled data. Second, the proposed social tracking method with MRF can further update the microblog event classification results towards a better performance.

4.4. Comparison with CR and CS

We further compare the proposed method with the state-of-the-art methods, i.e., CR [Reuter and Cimiano 2012] and CS [Becker et al. 2010]. As shown in the results, the proposed method significantly outperforms CR and CS in all evaluation measures. Especially, M^{2}DN+ST achieves an improvement of 47.0%, 51.0%, and 48.9% compared with CR on the recall, precision, and F, respectively. Compared with CS, the improvement is 57.2%, 60.7%, and 58.9%, respectively. This result can demonstrate that the proposed method is effective on relevant microblogs classification for events.

4.5. On M^{2}DN

Here, we evaluate the M^{2}DN method. We compare the experimental results between M^{2}DN and SVM, and between M^{2}DN+ST and SVM+ST. As shown in the results from

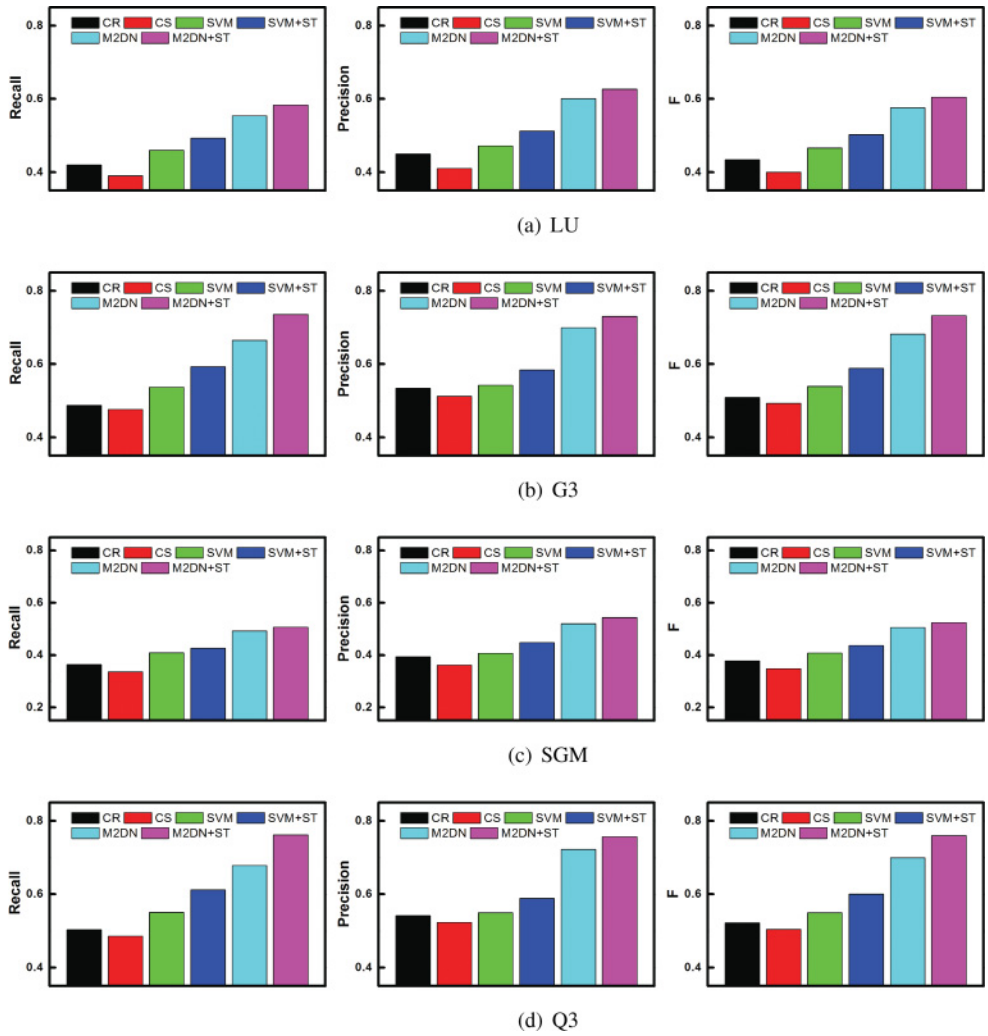

Fig. 5. Detailed results for four example events.

Figure 4, M^{2}DN outperforms SVM with an improvement of 23.8%, 29.0%, and 26.5% with respect to Recall, Precision, and F, respectively. We have similar observations for the comparison between M^{2}DN+ST and SVM+ST, where M^{2}DN+ST outperforms SVM+ST with 24.0%, 25.1%, and 24.5% with respect to Recall, Precision, and F, respectively. These results can demonstrate that M^{2}DN is effective on microblogs prediction. M^{2}DN benefits from its new two-paths deep learning architecture, which is able to handle the multi-modal multi-instance weakly-labeled classification task. The incompatible meaning problem between the text and the corresponding image for microblogs can be ameliorated by using the proposed method. It is noted that M^{2}DN does not require the accurate annotation on the text and the visual content for a given event. The well-designed deep learning framework can extract the underneath representation for event information automatically. Therefore, the proposed method can achieve a better performance in comparison with the state-of-the-art methods even without the further social tracking with MRF procedure.

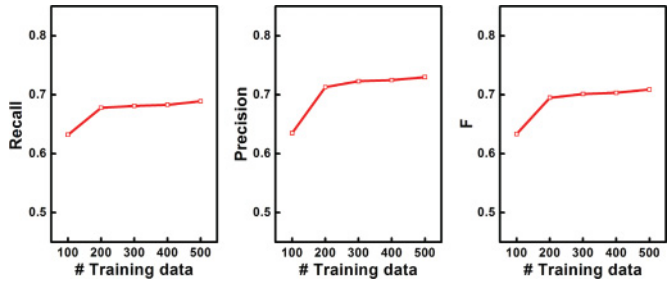

Fig. 6. The average performance curves with respect to the variation of training data.

4.6. On the Social Tracking with MRF

Here, we evaluate the proposed social tracking with MRF method by the comparison between M^{2}DN+ST and M^{2}DN, and the comparison between SVM+ST and SVM. As shown in the results of the 8th week, M^{2}DN+ST outperforms M^{2}DN with an improvement of 10.5%, 4.4%, and 7.5% with respect to Recall, Precision, and F, respectively. For the SVM method, the improvement by ST is larger. SVM+ST outperforms SVM with an improvement of 10.3%, 7.8%, and 9.4% with respect to Recall, Precision, and F, respectively. These results can justify the effectiveness of the proposed ST on microblogs relevance prediction. The proposed ST can find more relevant microblogs for the testing sample, which can help to enrich the testing microblogs. The MRF modeling further employs these data as the auxiliary information for microblogs prediction from the social connection view. As each user is not isolated in social network and most microblogs are not alone, the proposed method can explore the social connection among microblogs, which can provide more background information for the testing data. The better performance can be attributed to the following reason. The proposed method benefits from the use of auxiliary social information for microblogs event classification. Generally, each microblogs is very short and conveys little information. With the social tracking, we can obtain more social-related microblogs, which are the surrounding information for the testing data. Based on the textual and visual content analysis, these microblogs are employed in the MRF model to update the single microblogs event classification results. Therefore, the limited information problem can be tackled in the proposed method, which can lead to about 5% to 10% performance improvement, as shown in the results. These results can demonstrate the effectiveness of the social tracking algorithm.

4.7. On the Training Data

Here, we evaluate the influence of training data for M^{2}DN. We vary the training data with respect to the number of positive samples for each event. It is noted that the employed negative samples for training keep at about three times that of the positive samples. The experimental results are shown in Figure 6. As shown in the results, the increasing training data can lead to better event classification results, and generally, the proposed method can achieve a steady performance after the number of positive samples is larger than 200. Therefore, we employ 200 positive samples for each event in the training procedure.

5. CONCLUSION

In this article, we employ a deep learning architecture, named M^{2}DN, with social tracking on MRF for event classification in microblogs. In our method, a two-path deep network is trained to ameliorate the influence of weakly labeled data coming from the

incompatible meanings between the text and images of microblogs. Then, the social information is employed to find relevant microblogs as auxiliary information to enrich the testing sample using the MRF model. In this work, we evaluate the proposed method on the Brand-Social-Net dataset for event classification.

We can conclude from experimental results that: first, the M^2DN learning framework is effective on microblogs classification, and second, the proposed social tracking with MRF is able to locate relevant microblogs very efficiently to enhance the M^2DN results. The proposed method can achieve a high recall in comparison with the state-of-the-art methods.

In this work, we only take the event classification task as an example, and the proposed method can be extended to other types of entities. It is noted that there are still many future works. First, the entities in the social media platforms evolve very fast. Besides prediction of existing entities, it is important to automatically learn new entities from large-scale social media streams. Second, the use of social information in microblogs analysis should be deeply investigated. The society information behind the users can reveal the underneath property of the user and the corresponding social posts. The current work has some limitations. First, the training of M^2DN requires sufficient data. It is important to further explore other related resources to obtain the required data. Second, social media data always appears in a cross-platform way. Current work mainly focuses on single platform, which canbe further extended to support cross-platform analysis.

REFERENCES

J. Allan, R. Papka, and V. Lavrenko. 1998. On-line new event detection and tracking. In *Proceedings of the Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*.

E. Amigó, J. Gonzalo, J. Artiles, and F. Verdejo. 2009. A comparison of extrinsic clustering evaluation metrics based on formal constraints. *Information Retrieval* 12, 4 (2009), 461–486.

Sitaram Asur and Bernardo A. Huberman. 2010. Predicting the future with social media. In *Proceedings of IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, Vol. 1. IEEE, 492–499.

Boris Babenko, Ming-Hsuan Yang, and Serge Belongie. 2009. Visual tracking with online multiple instance learning. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 983–990.

H. Becker, M. Naaman, and L. Gravano. 2010. Learning similarity metrics for event identification in social media. In *Proceedings of ACM International Conference on Web Search and Data Mining*. 291–300.

Yoshua Bengio. 2009. Learning deep architectures for AI. *Foundations and Trends in Machine Learning* (2009).

Jeff Donahue, Yangqing Jia, Oriol Vinyals, Judy Hoffman, Ning Zhang, Eric Tzeng, and Trevor Darrell. 2013. Decaf: A deep convolutional activation feature for generic visual recognition. *arXiv Preprint arXiv:1310.1531*

David Easley and Jon Kleinberg. 2010. *Networks, Crowds, and Markets*. Cambridge University Press.

Weiguo Fan and Michael D. Gordon. 2014. The power of social media analytics. *Commun. ACM* 57, 6 (2014), 74–81.

Yue Gao, Fanglin Wang, and Tat-Seng Chua Huabo Luan. 2014. Brand data gathering from live social media streams. In *Proceedings of ACM Conference on Multimedia Retrieval*.

Yue Gao, Yi Zhen, Haojie Li, and Tat-Seng Chua. 2016. Filtering of brand-related microblogs using social-smooth multiview embedding. *IEEE Transactions on Multimedia* (2016).

David Hardoon, Sandor Szedmak, and John Shawe-Taylor. 2004. Canonical correlation analysis: An overview with application to learning methods. *Neural Computation* 16, 12 (2004), 2639–2664.

Qi He, Kuiyu Chang, and Ee-Peng Lim. 2007. Analyzing feature trajectories for event detection. In *Proceedings of the Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. 207–214.

M. Hearst. 2009. Search User Interfaces. Cambridge University Press.

Alexander T. Ihler, John W. Fisher III, Alan S. Willsky, and David Maxwell Chickering. 2005. Loopy belief propagation: Convergence and effects of message errors. *Journal of Machine Learning Research* 6, 5 (2005).

Andreas M. Kaplan and Michael Haenlein. 2010. Users of the world, unite! The challenges and opportunities of social media. *Business Horizons* 53, 1 (2010), 59–68.

Andrej Karpathy, George Toderici, Sachin Shetty, Tommy Leung, Rahul Sukthankar, and Li Fei-Fei. 2014. Large-scale video classification with convolutional neural networks. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 1725–1732.

Koray Kavukcuoglu, Pierre Sermanet, Y-Lan Boureau, Karol Gregor, Michaël Mathieu, and Yann L. Cun. 2010. Learning convolutional feature hierarchies for visual recognition. In *Proceedings of Advances in Neural Information Processing Systems*. 1090–1098.

Jon Kleinberg. 2002. Bursty and hierarchical structure in streams. In *Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 91–101.

Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. 2012a. Imagenet classification with deep convolutional neural networks. In *Proceedings of Advances in Neural Information Processing Systems*. 1097–1105.

Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. 2012b. ImageNet classification with deep convolutional neural networks. In *Proceedings of Advances in Neural Information Processing Systems*.

Svetlana Lazebnik, Cordelia Schmid, and Jean Ponce. 2006. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*, Vol. 2. IEEE, 2169–2178.

C. D. Manning, P. Raghavan, and H. Schütze. 2008. Introduction to information retrieval. In *Cambridge Univ. Press*.

Oded Maron and Aparna Lakshmi Ratan. 1998. Multiple-instance learning for natural scene classification. In *Proceedings of International Conference on Machine Learning*, Vol. 98. Citeseer, 341–349.

Megha Pandey and Svetlana Lazebnik. 2011. Scene recognition and weakly supervised object localization with deformable part-based models. In *Proceedings of IEEE International Conference on Computer Vision*. IEEE, 1307–1314.

Gabriel Pui, Cheong Fung, Jeffrey Xu Yu, Philip S. Yu, and Hongjun Lu. 2005. Parameter free bursty events detection in text streams. In *Proceedings of the International Conference on Very Large Data Bases*. 181–192.

Ning Qian. 1999. On the momentum term in gradient descent learning algorithms. *Neural Networks* 12, 1 (1999), 145–151.

Timo Reuter and Philipp Cimiano. 2012. Event-based classification of social media streams. In *Proceedings of the 2nd ACM International Conference on Multimedia Retrieval*.

Hassan Sayyadi, Matthew Hurst, and Alexey Maykov. 2009. Event detection and tracking in social streams. In *Proceedings of ACM International Conference on Web Search and Data Mining*.

Bharath Sriram, Dave Fuhry, Engin Demir, Hakan Ferhatosmanoglu, and Murat Demirbas. 2010. Short text classification in Twitter to improve information filtering. In *Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, 841–842.

A. Strehl, J. Ghosh, and C. Cardie. 2002. Cluster ensembles - a knowledge reuse framework for combining multiple partitions. *Journal of Machine Learning Research* 3 (2002), 583–617.

Bruce Thompson. 2005. Canonical correlation analysis. *Encyclopedia of Statistics in Behavioral Science.*

Tijmen Tieleman. 2008. Training restricted Boltzmann machines using approximations to the likelihood gradient. In *Proceedings of International Conference on Machine Learning*.

Jian Shu Weng and Bu Sung Lee. 2011. Event detection in Twitter. In *Proceedings of International Conference on Weblogs and Social Media*.

Y. Yang, T. Pierce, and J. G. Carbonell. 1998. A study on retrospective and on-line event detection. In *Proceedings of the Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*.

Xiangmin Zhou and Lei Chen. 2014. Event detection over twitter social media streams. *International Journal on Very Large Data Bases* 23, 3 (2014), 381–400.

Received May 2015; revised June 2016; accepted June 2016