ROBUST LATENT POISSON DECONVOLUTION FROM MULTIPLE IMPERFECT FEATURES FOR WEB TOPIC DETECTION

Fei Tao[⋆]*, Junbiao Pang*[‡]*, Chunjie Zhang*[⋆†]*, Liang Li*[⋆†]*, Li Su*[⋆†]*, Weigang Zhang*[*]*, Qingming Huang*[⋆†]*, Guiping Su*[⋆]

[⋆]School of Computer and Control Engineering, University of Chinese Academy of Sciences, Beijing, China
[‡]Beijing Key Laboratory of Multimedia and Intelligent Software Technology,
College of Metropolitan Transportation, Beijing University of Technology, China
[†]Key Lab on Big Data Mining and Knowledge Management, Chinese Academy of Sciences, Beijing, China
[*] School of Computer Science and Technology, Harbin Institute of Technology, China
fei.tao@vipl.ict.ac.cn {jbpang, cjzhang, lli, wgzhang, qmhuang}@jdl.ac.cn {suli, sugp}@ucas.ac.cn

ABSTRACT

In web topic detection, detecting "hot" topics from enormous User-Generated Content (UGC) on web data poses two main difficulties that conventional approaches can barely handle: 1) poor feature representations from noisy images and short texts; and 2) uncertain roles of modalities where visual content is either highly or weakly relevant to textual cues due to less-constrained data. In this paper, following the detection by ranking approach, we address the problem by learning a robust shared representation from multiple, noisy and complementary features, and integrating both textual and visual graphs into a k-Nearest Neighbor Similarity Graph (k-N^{2}SG). Then Non-negative Matrix Factorization using Random walk (NMFR) is introduced to generate topic candidates. An efficient fusion of multiple graphs is then done by a Latent Poisson Deconvolution (LPD) which consists of a poisson deconvolution with sparse basis similarities for each edge. Experiments show significantly improved accuracy of the proposed approach in comparison with the state-of-the-art methods on two public data sets.

Index Terms— Similarity Cascade, Latent Poisson Deconvolution, Multi-view Learning, Topic Detection, Cross Media

1. INTRODUCTION

With the rapid development of social media, User-Generated Content (UGC) [1] is quite pervasive for people to either share or exchange their options and experiences. Meanwhile, the content of UGC is more sparse, unconstrained and less predicable than that of the professionally edited articles since that everybody is both the producer and the consumer of media. As a result, the unprecedented explosion in the volume of "we-media" has made it difficult for web users to quickly access hot and interesting topics [2]. Web topic detection [1] is such an effort to detect and organize web data into meaningful topics automatically.

One of important approaches is to exploit multiple modalities of data itself. Generally speaking, "we-media" is often delivered at will by multiple modalities and reflects social realities from more aspects. Therefore, these less-constrained UGC data often faces several challenging problems: 1) the possible deficiency of modalities; 2) inefficient feature representations for short and noisy texts [3]; and 3) the uncertain roles of different modalities. The last is extremely universal phenomenon, where visual cues are more important than textual ones to express the content for some webpages; on the other side, textual cues serve as a dominant role for the other ones. To handle multiple modalities of data, the existing methods simply combine multiple representations of different modalities with possibly noise, which may often degrade the performance. For instance, different modalities are simply weighted into a unified representation by well-tuned weights [4]. However, the varying roles of different modalities are not considered.

Therefore, we seek a *robust*, *adaptive* and *data-level* framework to exploit these *multiple* and *imperfect* feature representations, based on two motivations. First, although an enormous volume of literature has been devoted to multiple features fusion, most of them consensually assume that features nearly have no noise. Second, we want to avoid the disadvantages of popular modality-level fusion methods, e.g., linear weights or non-linear ones. The modality-level methods obviously have a difficulty in handling with the uncertain role of different modalities, and also are sensitive to noise.

This work was supported in part by the National Basic Research Program of China (973 Program) under Grant 2012CB316400 and 2015CB351800, in part by the Natural Science Foundation of China under Grant 61332016, Grant 61202234, Grant 61303153, Grant 61402431, Grant 61303154, Grant 61472389 and Grant 61472387, in part by the Beijing Post-Doctoral Research Foundation, in part by College Students Innovation and Practice Training Program of CAS, and in part by the Funding Project for Academic Human Resources Development in Institutions of Higher Learning Under the Jurisdiction of Beijing Municipality.

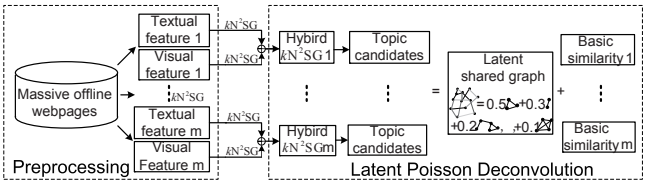


Fig. 1. Proposed Latent Poisson Deconvolution framework.

In summary, our goal is to robustly detect topics from multiple imperfect representations from multi-media data where the roles of modalities vary respect to each individual input.

In this paper, we propose a Latent Poisson Deconvolution (LPD) framework to explicitly handle the possible noise associated with different feature representations. As shown in Fig. 1, following Similarity Cascade (SC) [1], we firstly in the pre-processing stage extract multi-view features from different modalities, and compute a similarity graph with each feature representation. The adverse impacts of imperfect representations are also naturally encoded into similarities. To partially reduce these unfavorable impacts, we only select top-k similar values to generate k-Nearest Neighbor Similarity Graph (kN^{2}SG). Then, each paired multi-view kN^{2}SGs is equally weighted into a hybrid graph. Instead of directly using kN^{2}SG from each modality, the motivation of the hybrid graph is that visual cues still contain some minor yet meaningful information, although it is extremely difficult to convert visual cues into social-related concepts due to the "semantic gap" challenge.

Further in generating topic candidates on these hybrid graphs, Non-negative Matrix Factorization using Random walk (NMFR) [5] is used to generate multi-granularity candidates. The advantage lies in that random walk empowers NMF to capture manifold similarity.

In the detection by ranking stage [1], rather than firstly learning a latent shared graph (which is a common approach in multi-view learning, however unnecessary in our approach as will be discussed) we instead propose LPD to learn the reconstructed latent shared graph and sparse basic similarities from different hybrid graphs, in the hope of not only avoiding complex optimization problems but also adaptively fusing multiple imperfect cues at the data-level. The last is crucial to deal with the uncertain roles of modalities in the individual webpage.

2. RELATED WORK

Topic Detection from Multi-modalities: This approach aims to exploit the possible complementary modalities from data itself. One modality is often considered as the mutual side information of the others. In this approach, there are two important threads of research. One extends clustering algorithms into the multi-modality data [6] [7], and the other is the similarity graph method [8], a work based on multi-modalities fusion.

In the former case, discovery of topics involves extending single-modality based models into multi-modal data. For instance, Multi-modal LDA [7] was proposed to group image with tags into topics. In similarity graph method, multi-modal information is fused into edges of a similarity graph. For instance, In [9], Wu et al. used weighted similarity between Nearly-Duplicated Keyframe (NDK) and text based on speech transcripts for news videos. Compared with the extension of topic modelings [10] into multi-modal data, similarity graph is computationally simple, and is easily extendable for other graph-based algorithms [11]. In contrast, this paper does not necessarily aim at designing a perfect weight scheme to fuse heterogenous graphs, but rather adaptively fuses multiple similarities at each edge.

Multi-view Learning: Multi-view learning is the problem of machine learning from data represented by multiple feature sets. In this paper, multiple views mean heterogenous descriptions of a given sample. Many methods have been proposed for multi-view classification [12], retrieval [13], clustering [14]. We refer the interested readers to [15] for a literature survey of multi-view learning. To the best of our knowledge, we firstly apply multi-view learning to unsupervised topic detection.

Technically, in our multi-view learning, each view corrupted by noise is explicitly handled. Our approach treats each view as corrupted by a sparse noise and learns a latent shared graph by exploiting the joint view statistics.

3. GENERATING CANDIDATES ON HYBRID GRAPH

3.1. Combining kN^{2}SGs into a Hybrid Graph

Given a set of data points $\{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$ where each sample $\mathbf{x}_i = (x_i^v, x_i^t)$ contains the visual modality x_i^v and the textual one x_i^t. In the context of multi-view learning, we are given a set of data points in M views. In each of view we can construct a pair-wise similarity matrix S_v^m and S_t^m for visual and textual modalities respectively. Let $s_{ij} \geq 0$ denotes the similarity on a pair of data point $\mathbf{x}_i$ and $\mathbf{x}_j$[1].

Further, let (V, E, W^m) be a weighted graph with vertex set V, edge set E, and the corresponding weight/similarity matrix S^m, where each vertex v_i associates with the webpage $\mathbf{x}_i$ and edge $(i, j) \in E$ associate with $W^m(i, j)$ between $\mathbf{x}_i$ and $\mathbf{x}_j$. We propose to use Gaussian kernels to covert the similarity matrix S^m, i.e., $w_{ij}^m = \exp\left(-\|s_{ij}^m\|^2/\sigma^2\right)$ where $\|\cdot\|^2$ denotes the ℓ_2 norm and σ^2 denotes the deviation. Gaussian kernel nonlinearly scales different similarity S_m into a uniformly one ranging from 0 to 1. Because different features

[1]In the following, we ignore the superscript and subscript in the different context, if it does not cause confusion.

from diverse modalities not necessarily have the same similarity measurement.

Once similarity graphs are computed, the top-k most similar data $\mathbf{x}_i$ are inserted as its neighbors on the graphs, and the other similarities are assigned with zeros. Consequently, some unnecessary correlations among data are simply filtered out. We term the resulting sparse graphs as k-Nearest Neighbor Similarity Graph (kN^{2}SG). Subsequently, a pair of visual kN^{2}SG (V, E_v^m, W_v^m) and the textual one (V, E_t^m, W_t^m), from some view of data $\mathbf{x}_i$, are equally merged into a hybrid graph, $\mathcal{G}^m = (V, E_v^m \cup E_t^m, (W_v^m + W_t^m)/2)$.

3.2. Generating Topic Candidates

NMFR is carefully chosen to generate topic candidates, in order to exploit the non-neighborhood relationship on these sparse graphs G^m. Let $U^m \in \mathbb{R}^{N\times K}$ be nonnegative and orthogonality matrix, the objective function of NMFR is as:

$$\begin{aligned}\min_{U^m\geq 0} \; & -\mathrm{Tr}({U^m}^\top A U^m) + \lambda\|U^m\|_F^2 \\ s.t. \; & {U^m}^\top U^m = I.\end{aligned}$$

A is random walk distance, $A = (I - \alpha D^{m^{-1/2}} G^m D^{m^{-1/2}})^{-1}$, where $\alpha \in (0,1)$ is a decay parameter and D^m is a diagonal matrix with $D^m(i,i) = \sum_j W^m(i,j)$. [5] proposes a relaxed algorithm to optimize U^m without explicitly computing matrix A.

By the winner-take-all principle, U^m generates the topic candidates C_k^m $(k = 1, \ldots, K)$. Formally, $C_k^m = {b_k^m}^\top \circ b_k^m$ where the indicator vector $b_k^m \in \{0,1\}^N$, in each of whose bin 1 or 0 means that topic C_k^m whether contains the $\mathbf{x}_i$ or not. The operation $\circ$ means that the diagonal of matrix ${b_k^m}^\top b_k^m$ is set to zero.

4. ROBUST LATENT POISSON DECONVOLUTION

4.1. Latent Poisson Deconvolution

By our notation for hybrid graph G^m, the basic assumptions in LPD are threefolds: 1) The graph G^m in each individual hybrid view is sufficient to discover most of correlations; 2) The values in each G^m are corrupted by noise; and 3) Different G^m has different basic similarities to indicate the "absolute" correlations. Based on these assumptions, the similarity w_{ij}^m can be naturally decomposed into three parts as follows:

$$\forall m, \;\; G^m = G + B^m + noise, \tag{1}$$

where G is a shared latent graph that reflects the underlying true correlation among webpages, combined with different basic similarities B^m and a noise term. Ideally, once G is computed, we can simply use G as the input of Poisson Deconvolution [1] to rank topics. A key question arising here is how to model the latent graph G, the basic similarity b_{ij}^m and $noise$ term.

Following Poisson assumption in PD [1], $noise$ is the Poisson distribution to model the long-tailed noise. Therefore, optimizing G in Poisson noise involves the derivation of the function $\ln(G_{ij}!)$ with respect to each edge, resulting in an inefficient numerical solution. More efficiently, we can approximate $G \approx \sum_{k=1}^K \mu_k C_k$ where $C_k = \mathrm{union}(C_k^m), m = 1, \ldots, M$, and μ_k are the ranking scores of topics C_k. Each basic similarity B^m represents the difference between G and G^m. Since we assume the graph G^m in each hybrid view are sufficient to identify most of the shared graph G, it is reasonable to assume that only a small fraction of elements in G^m being significantly different from the corresponding ones in G. That is, B^m tends to be sparse.

Under the above assumptions, we formulate LPD to rank topics as the problem of the minimization of the negative log-likelihood:

$$\begin{aligned}\min \mathcal{L}(\mu_k, B^m) = -\ln \prod_{m=1}^M \prod_{w_{ij}^m \in G^m} \frac{{w_{ij}^m}^{G_{ij}^m} e^{-w_{ij}^m}}{G_{ij}^m!} + \lambda \sum_{i=1}^M \|B^m\|_1, \\ s.t. \;\; \mu_k \geq 0, \;\; k = 1, \ldots, K\end{aligned} \tag{2}$$

where λ is a non-negative trade-off parameter, and the ℓ_1 norm $\|B^m\|_1$ is well-known to be a convex non-smooth function. The constraints $\mu_k \geq 0$ lead to a nearly sparse solution, automatically removing the redundancy of the topic candidates. The first term in (2) learns a reconstructed, shared and latent graph from M hybrid ones.

4.2. Optimization

The optimization problem in (2) is still challenging due to its non-differentiable and non-smooth terms $\|B^m\|_1$. We apply the idea of Alternating Direction Method of Multipliers(ADMM) [16] to convert the optimization problem to several sub-problems by introducing auxiliary variables Z^m, $m = 1, \ldots, M$

$$\begin{aligned}\min_{\boldsymbol{\mu}\geq 0, B^m, Z^m} \; & \mathcal{L}(\boldsymbol{\mu}, B^m) + \lambda \sum_{m=1}^M \|B^m\|_1 \\ s.t. : \; & \boldsymbol{\mu} \geq 0, B^m = Z^m, \;\; m = 1, \ldots, M\end{aligned} \tag{3}$$

In ADMM, we optimize the augmented Lagrangian of the above problem that can be formulated as follows:

$$\begin{aligned}\mathcal{L}_\rho =& \mathcal{L}(\boldsymbol{\mu}, B^m) + \lambda \sum_{m=1}^M \|Z^m\|_1 + \frac{\rho}{2} \sum_{m=1}^M \|B^m - Z^m\|_F^2 \\ & + \sum_{m=1}^M \mathrm{trace}({Y^m}^\top (B^m - Z^m)),\end{aligned} \tag{4}$$

where $\rho \geq 0$ is called the penalty parameter and $\|\cdot\|_F$ denotes the Frobenius norm, the matrices Y^m are the dual variables associated with the constraints $B^m = Z^m$, respectively. The algorithm for solving the above augmented Lagrangian problem involves the following iterative steps:

$$\begin{aligned}\boldsymbol{\mu}^{t+1}, {B^m}^{t+1} &= \arg\min_{\boldsymbol{\mu}\geq 0, B^m} \mathcal{L}_\rho(\boldsymbol{\mu}, B^m, {Z^m}^t, {Y^m}^t) \\ {Z^m}^{t+1} &= \arg\min_{Z^m} L_\rho(\boldsymbol{\mu}^{t+1}, {B^m}^{t+1}, Z^m, {Y^m}^t) \\ {Y^m}^{t+1} &= {Y^m}^t + \rho({B^m}^{t+1} - {Z^m}^{t+1})\end{aligned} \tag{5}$$

The advantage of sequential update is that we separate multiple variables and thus can optimize them at a time.

Solving for $\boldsymbol{\mu}$ and B^m: When solving for $\boldsymbol{\mu}$ and B^m in (4), the relevant terms from $\mathcal{L}_\rho$ are

$$\begin{aligned}\arg \min_{\boldsymbol{\mu}\geq 0, B^m} \mathcal{L}(\boldsymbol{\mu}, b^m) + \sum_{m=1}^{M} \text{trace}({Y^m}^\top (B^m - Z^m)) \\ + \frac{\rho}{2} \sum_{m=1}^{M} \|B^m - Z^m\|_F^2\end{aligned} \tag{6}$$

By Jansen's inequality, (6) has the following upper bound:

$$\begin{aligned}-\sum_{m=1}^{M} \sum_{w_{ij}^m \in G^m} (G_{ij}^m (\sum_{k=1}^{K} p_k \ln \frac{\mu_k C_k}{p_k} + P_{ij}^m \ln \frac{B_{ij}^m}{P_{ij}^m}) - \sum_{k=1}^{K} \mu_k C_k \\ -B_{ij}^m) + \sum_{m=1}^{M} \text{trace}({Y^m}^\top (B^m - Z^m)) + \frac{\rho}{2} \sum_{m=1}^{M} \|B^m - Z^m\|_F^2,\end{aligned} \tag{7}$$

where p_k and P_{ij}^m are the hidden variables that satisfy $P_{ij}^m + \sum_{k=1}^{K} p_k = 1$ ($P_{ij}^m \geq 0$, $p_k \geq 0$), $p_k = \frac{\mu_k C_k}{B_{ij}^m + \sum_{k=1}^{K} \mu_k C_k}$, and $P_{ij}^m = \frac{B_{ij}^m}{B_{ij}^m + \sum_{k=1}^{K} \mu_k C_k}$.

Solving (6) by minimizing the upperbound of (7) has the closed form solutions, and the non-negativity constraints are automatically taken care of:

$${B_{ij}^m}^{t+1} = (-A + \sqrt{A^2 + 4\rho C})/2\rho \tag{8}$$

$$\mu_k^{t+1} = \frac{\sum_{m=1}^{M} \sum_{(i,j)\in G^m} G_{ij}^m p_k}{M \sum_{(i,j)\in G^m} C_k} \tag{9}$$

where

$$A = Y^m - \rho Z^m + 1_{(i,j)\in G^m}, \;\; C = G_{ij}^m P_{ij}^m$$

Solving for Z^m: The optimization problem for Z^m can be equivalently written as

$$\min_{Z^m} \lambda \|Z^m\|_1 + \frac{\rho}{2} \|Z^m - Y^m/\rho - B^m\|_F^2,$$

which has a closed form solution:

$${Z^m}^{t+1} = \mathcal{S}_{\lambda/\rho}(Y^m/\rho + B^m), \tag{10}$$

where $\mathcal{S}_\alpha(\mathbf{x}) = \max(\mathbf{x} - \alpha, 0) + \min(\mathbf{x} + \alpha, 0)$ is the shrinkage operator [17].

Since the objective (2) is convex subject to linear constraints, and all of its subproblems can be solved exactly, based on existing theoretical results [18] LPD converges to global optima.

Once μ_k and B^m are computed, the interestingness of topics are computed as, $i_k = \mu_k \cdot |C_k|$, where $|C_k|$ is the number of objects in topic C_k [1]. Note that during the evaluation, our method adopts the Non-Maximal Suppression (NMS) [19] to handle the problem that which one is selected as the real topic if several topics intersect with each others.

5. EXPERIMENT AND DISCUSSION

We evaluate our method on two public data sets, i.e., MCG-WEBV [20] and YKS [4]. MCG-WEBV is downloaded from the "Most viewed" videos of "This month" on YouTube. For MCG-WEBV, the surrounding text of each video is considered as a set of words. While YKS is a cross-media data set crawled from YouKu and Sina respectively.

We chose two multi-view features for textual cues, i.e., Latent Dirichlet Allocation (LDA) [10] and TF-IDF, and Fisher Vector (FV) [21] to describe images. In our experiments, the dictionary size of LDA is 1,000. FV with 256 Gaussian components is used to represent keyframes in a video clip, where SIFT points are densely sampled from 24×24 patches. Once keyframes are encoded, video signature [22] is computed as similarity between two clips. The cosine distance is used to measure the similarity between textual features. Therefore, two hybrid graphs, $TF-IDF$+FV and LDA+FV, are generated. k is assigned 100 for MCG-WEBV and 20 for YKS in textual features. And for video graphs, k is all assigned 5 for both datasets.

In all the experiments, we use two metrics to measure the performances: Top-10 F_1 versus Number of Detected Topics (NDT) and accuracy versus False Positive Per Topic (FPPT) [1]. Note that if two methods have the same top-10 F_1 or accuracy score, the one with smaller NDT or FPPT has better performance for both metrics.

In all the experiments, the SC with thresholds of $[0.1, 0.5, 0.9]$ and in detecting candidates using NMFR, the numbers of clusters are $[100, 500, 900, 1300]$ and the random walk extent λ = 0.8.

5.1. Analysis of Our Approach

5.1.1. The Analysis of k-N^{2}SGs

In order to show that k-N^{2}SGs can partially remove noise in similarity graph and further achieve better performance, 100-N^{2}SGs and Full Similarity Graphs (FSGs) are separatively built from hybrid graphs of TF-IDF+FV and LDA+FV. Then, NMFR is applied on these graphs to generate 4,240 and 3,445 candidates from 100-N^{2}SGs and 4,245 and 3,289 ones from FSGs, respectively. Accuracy versus FPPT is used to evaluate the performance.

As shown in Fig. 2, 100-N^{2}SG achieves a higher accuracy than that of FSGs when FPPT is smaller than 6. Also noticed accuracies output by 100-N^{2}SG increase faster, and as a result the 100-N^{2}SG outperforms FSG about 25% accuracy.

5.1.2. The Effectiveness of Latent Shared Graph Approach

Fig. 3(a) illustrates the effectiveness of single modality and multi-modalities. The results of LPD largely outperform that of LDA+FV. Compared with the results from $TFIDF$+FV, LPD obtains very similar results when FPPT

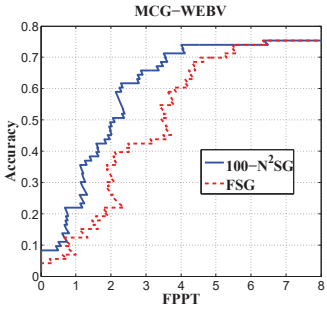
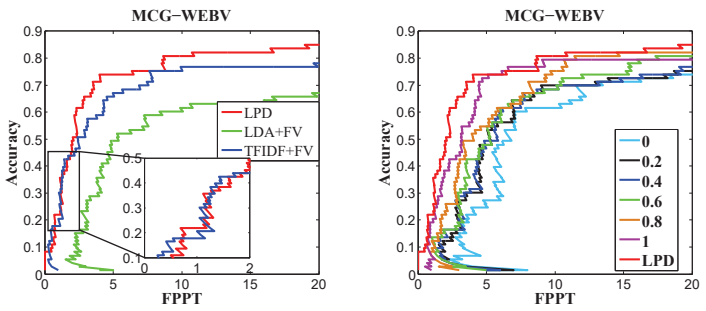


Fig. 2. k-N^{2}SG versus FSG.



(a) Single modality versus LPD

(b) Linear weighting scheme (11) versus LPD

Fig. 3. Comparisons among single modality, linear weighting scheme and LPD.

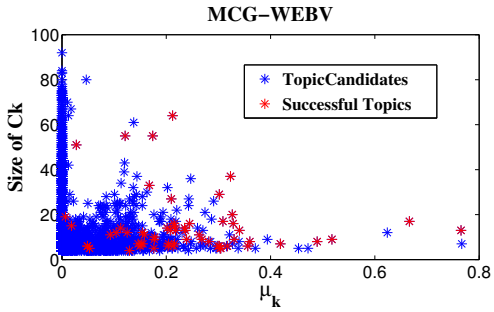
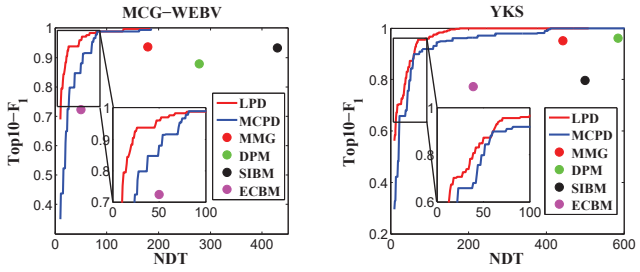


Fig. 4. Relationship between μ_k and $|C_k|$. Blue points are the topic candidates generated by k-N^{2}SG, red ones are the successfully detected topics (best viewed in color).



(a) (b)

Fig. 5. Comparisons between the state-of-the-art methods and our method by top-10 F_1 versus NDT on two datasets (best viewed in color).

is less than 2, but surpasses it by about 5% accuracies where FPPTs are from 3 to 20. The consistently improved results show that the latent shared graph can efficiently fuse the complementary information from multi-hybrid graphs.

A conventional method for multi-feature fusion assumes that every modality or feature has a linear nonnegative weight:

$$G_{fusion} = \alpha * G_{TFIDF+FV} + (1-\alpha) * G_{LDA+FV}, \qquad (11)$$

where α is a weight. In our experiments, a set of α, $\{0, 0.2, 0.4, 0.6, 0.8, 1.0\}$, is used to tune the best fusion parameter; then, the resulting G_{fusion} is deconvoluted by PD method [1]. Fig. 3(b) illustrates the comparisons between our method and the linear weight scheme. As is expected, the linear combination of $TFIDF$+FV and LDA+FV does not obtain improved results. Interestingly LDA+FV even drags down the performances of $TFIDF$+FV. On the contrary, LPD consistently improves the performances for both graphs. These results indicate that the data-level and modality-sensitive fusion approach is more suitable for UGC data.

5.1.3. Analysis of the Relationship Between μ_k and $|C_k|$

Fig. 4 illustrates two important aspects from the relationship between μ_k and $|C_k|$. The first is about 85% μ_k are nearly equal to zeros ($\leq 1e^{-4}$); besides, the size of these zero-weight topics ranges from 4 to 90. This indicates that LPD is quite robust to the number of candidates because only a few meaningful ones are selected. The second is that there is no close correlation between μ_k and the size of C_k. These observations in our experiments also illustrate the importance of the combination of the size of topic and its weight in identifying a real topic.

5.2. Comparisons With State-Of-The-Art Algorithms

In this section, we compare our method with Event-Clustering Based Method (ECBM) [20], Discriminative Probabilistic Models (DPM) [23] Maximal Cliques with Poisson Deconvoluton (MCPD) [1], Side-Information Based Method (SIBM) [24] and Multi-Modality Based Method (MMG) [4]. ECBM, DPM and MCPD are only based on textual modality, and MCPD achieves the state-of-the-art performances on both MCG-WEBV and YKS data sets. SIBM and MMG are based on textural and visual modalities.

Fig. 5 shows comparison results by Top-10 F_1 versus NDT. Our method achieves a higher Top-10 F_1 score than that of the others. Top-10 F_1 of our method increases quickly along with number of generated topics. For instance,

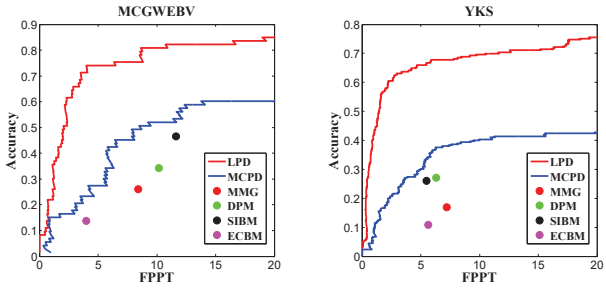


Fig. 6. Comparisons between the state-of-the-art methods and our method by FPPT versus Accuracy on two datasets (best viewed in color).

to achieve approximate 0.9 top-10 F_1 score, MCPD [1], MMG [4], SIBM [24] and DPM [23] generated 70, 179, 430, and 275 topics respectively on MCG-WEBV, while our method only requires 20 topic candidates. [zoom the details of results]

We also use accuracy versus FPPT to measure the performances of different algorithms. The accuracy versus FPPT can depict the relations between the number of correctly detected topics and the false positive rate of an algorithm. As shown in Fig. 6, our approach ("LPD") is consistently better than MMG [4], DPM [23], SIBM [24] and ECBM [20]. Our system significantly outperforms all the other state-of-the-art methods. For instance, at FPPT=5, our method achieves 0.78 and 0.67 accuracies on MCG-WEBV and YKS respectively. In contrast, MCPD only obtains 0.32 and 0.33 accuracies on MCG-WEBV and YKS, respectively.

6. CONCLUSION

In this paper, we have described a topic detection method based on fusing multiple imperfect features into a latent shared graph, leading to results matching or surpassing the state-of-the-art methods in web topic detection. There are significant distinctions between the proposed LPD and previous studies in fusing multi-modalities for topic detection:

(1) We demonstrated the effectiveness of adaptive, data-level feature fusion for topic detection and the effectiveness of topic detection by the latent shared representation.

(2) The proposed LPD enjoys the advantage of Poisson deconvolution method [1] for single modality in achieving high performance and that of the multi-view learning in its complementary information.

(3) The data-level adaptive fusion assumes no prior information about features, except the assumption of sparsity in basic similarity, in contrast to conventional linear weighting methods, which make the assumption that each modality is controlled by a weight.

7. ACKNOWLEDGE

The authors would like to thank Yucheng Sun with BJUT who helped us pre-process the datasets.

8. REFERENCES

[1] J. Pang, F. Jia, C. Zhang, W. Zhang, Q. Huang, and B. Yin, "Unsupervised web topic detection using a ranked clustering-like pattern across similarity cascades," *IEEE Trans. on MultiMedia*, vol. 17, no. 6, pp. 843–853, 2015.

[2] D.Shahaf and C.Guestrin, "Connecting the dots between news articles," in *SIGKDD*. ACM, 2010, pp. 623–632.

[3] W. X. Zhao, J. Jiang, J. Weng, J. He, E.-P. Lim, H. Yan, and X. Li, "Comparing twiter and traditional media using topic models," in *ECIR*, 2011, pp. 338–349.

[4] Y. Zhang, G. Li, L. Chu, S. Wang, W. Zhang, and Q. Huang, "Cross-media topic detection: a multi-modality fusion framework," in *ICME*, 2013, pp. 1–6.

[5] Z. Yang, T. Hao, O. Dikmen, X. Chen, and E. Oja, "Clustering by nonnegative matrix factorization using graph random walk," in *NIPS*, 2012, pp. 1079–1087.

[6] D. Blei and J. Lafferty, "A correlated topic model of science," *Annals of Applied Sciences*, vol. 1, pp. 17–35, 2007.

[7] D. Putthividhy, H.T. Attias, and S.S. Magarajan, "Topic regression multi-modal latent dirichlet allocation for image annotation," in *CVPR*, 2010, vol. 1, pp. 3408–3415.

[8] S. Papadopoulous, C. Zigkolis, Y. Kompatsiaris, and A. Vakali, "Cluster-based landmark and event detection on tagged photo collections," *IEEE Multimedia*, vol. 18, no. 1, pp. 52–63, 2011.

[9] X. Wu, G. Hauptmann, and C. Ngo, "Novelty detection for crosslingual news story with visual duplicates and speech transcripts," in *ACM Multimedia*, 2007, pp. 168–177.

[10] D. Blei, M. David, A. Ng, M. Jordan, and J. Lafferty, "Latent dirichlet allocation," *Journal of machine learning research*, vol. 3, pp. 993–1022, 2003.

[11] M. Aiello, G. Petkos, C. Martin, D. Corney, S. Papadopoulos, R. Skraba, A. Goker, I. Kompatsiaris, and A.Jaimers, "Sensing trending topics in twitters," *IEEE Trans.PAMI*, vol. 15, no. 6, pp. 1268–1282, 2013.

[12] A. Zien and C.S. Ong, "Multiclass multiple kernel learning," in *ICML*, 2007, pp. 1191–1198.

[13] B.Long, P.S.Yu, and Z.Zhang, "A general model for multiple view unsupervised learning," in *SDM*. 2008, pp. 147–159, Springer.

[14] S. Bickel and T. Scheffer, "Multi-view clustering," in *ICDM*, 2004, pp. 19–26.

[15] C. Xu, D. Tao, and C. Xu, "A survey on multi-view learning," arXiv preprint arXiv:1304.5634, 2013.

[16] S.Boyd, N.Parikh, E.Chu, B.Peleato, and J.Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.

[17] Z. Lin, M. Chen, and Y. Ma, "The augmented lagrage multipler method for exact recovery of corrupted low-rank matrices," arXiv preprint arXiv:1009.5055, 2010.

[18] Z.Q. Luo, "On the linear convergence of the alternating direction method of multipliers," arXiv preprint arXiv:1208.3922, 2010.

[19] M. Everingham, L. Van Gool, C. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes (voc) challenge," *IJCV*, , no. 3, pp. 303–338, 2010.

[20] J. Cao, C. Ngo, Y. Zhang, and J. Li, "Tracking web video topics: Discovery, visualization, and monitoring," *CSVT*, vol. 21, no. 12, pp. 1835–1846, 2011.

[21] J. Sánchez, T. Perronnin, T. Mensink, and J. Verbeek, "Image classification with the fisher vector: theory and practice," *IJCV*, vol. 105, no. 3, pp. 222–245, 2013.

[22] S. Cheung and A. Zakhor, "Efficient video similarity measurement with video signature," *CSVT*, vol. 13, no. 1, pp. 59–74, 2003.

[23] Q. He, K. Chang, E. Lim, and A. Banerjee, "Keep it simple with time: a reexamination of probabilitic topic detection models," *IEEE Trans.PAMI*, vol. 32, no. 10, pp. 1795–1808, 2010.

[24] T. Chen, C. Liu, and Q. Huang, "An effective multi-clue fusion approach for web video topic detection," in *ACM Multimedia*, 2012, pp. 781–784.