EVENT BASED NEWS VIDEO PEOPLE CLASSIFICATION AND RANKING USING MULTIMODALITY FEATURES

Chunxi Liu[1], Qingming Huang[1,2], Shuqiang Jiang[2],Changsheng Xu[3]
[1]Graduate University of Chinese Academy of Sciences, Beijing, 100190, China
[2]Key Lab of Intell. Info. Process., Inst. of Comput. Tech., CAS, Beijing, 100190, China
[3]Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China
{cxliu, sqjiang, qmhuang }@jdl.ac.cn, csxu@nlpr.ia.ac.cn

ABSTRACT

Existing research on news video analysis mainly concentrates on structure analysis, semantic concept detection, annotation and search. However, little work has been contributed to news video people community analysis, which is helpful for users to understand the event. In this paper, we propose a novel approach to classify the people appearing in the news video into different communities. In our approach, the people appearing in the news video are first identified by associating their faces with names. The faces are detected from the video frames, and the names are obtained from the text. Then, the people belonging to the same organization are clustered. After that, the relationships between these organizations are determined using sentiment analysis. The sentiment words are diverse in each news story and contain both positive and negative ones. However, we have news title, which is the summary of the story and the sentiment of which is clear, to help us to mine the relationships between the organizations. At last, social networks are built to classify those people/organizations into different classes, and the people/organizations are ranked in each community according to their influence. The main contributions of the paper are two folds: 1) we propose a novel approach to present the news video event according to communities; 2) we propose to use the sentiment analysis and social network to classify the news people/organizations. The experimental results on the selected news topics demonstrate that the proposed approach is effective.

Keywords— News video analysis, people classification and ranking, community mining, sentiment analysis, naming face, social network

1. INTRODUCTION

With the rapid development of digital communication technologies, personal computer capacities and WWW, more and more multimedia information is available in terms of online video data, personal video recordings, 24-hour br--oadcast news videos, etc. Among these data, news video has become an indispensable media in our daily life. However, due to the large volume of news video data, it is not easy for us to access our interested news topics. Therefore, there is a high demand for news video personalization. By using personalized service, we are able to optionally access and view our interested topics, which not only significantly save the time, but also enhance the power of current video content analysis system.

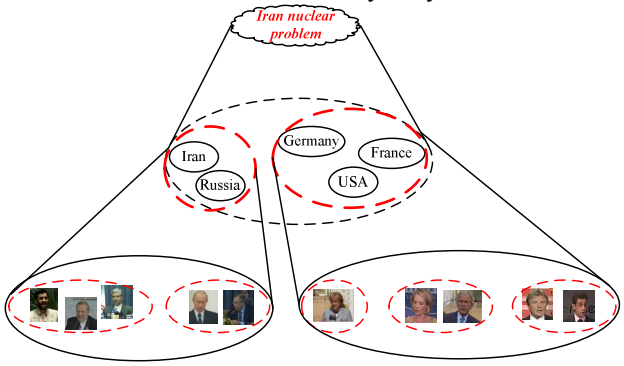

Fig. 1 An example of the news video people classification.

In order to facilitate people's accessing to news and searching video content, news video analysis has been a hot research topic for a long time. The state-of-the-art news video parsing covers structure analysis [1], semantic concept detection, annotation [2] and search [3]. Through structure analysis, the individual news story is segmented and news video can be accessed by story level indexing. After concept detection and annotation, the semantics in the videos can be understood. News video search can retrieve the user interested stories. Although much effort has been contributed to news video analysis, little work has been contributed to the news video people community analysis, which is helpful to understand the news event. News video is produced by collecting news from different sources and involves interviewing different people. For some critical events, different people/organizations may hold different perspectives. Those different attitudes toward the same event can be used to classify those people/organizations into different categories/communities. For example, for the controversial topic "*Iran nuclear problem*", there are obviously two communities with different opinions, one supports *Iran* while the other opposes *Iran*. The

This work was supported in part by National Natural Science Foundation of China: 60833006 and 60702035, and in part by Beijing Natural Science Foundation: 4092042.

classification example is shown in Fig. 1. Those people are classified into two communities. One community consists of *Putin*, *Lavrov*, *Hosseini*, *Mottati* and *Ahmadinejad*, while the other community consists of *Merkel, Perino*, *Bush, Kouchner* and *Sarkozy*. If those hidden communities in the news event can be discovered automatically from the news video, the users can browse the news event according to the different communities, and can understand the relationship between the people appearing in the news clearly.

In this paper we propose a novel approach to classify those people/organizations with different opinions in the critical news video event into different communities. The proposed framework is shown in Fig. 2. Firstly, the faces and name entities are extracted from the video and text respectively. Then, the names and faces are associated by naming face. After that sentiment analysis is used to analyze the relationship between the people/organizations. At last based on the built social networks, the people/organizations are classified into different communities and ranked according to their influence. The contributions of the paper include: 1) we propose a novel approach to present the news video event according to communities; 2) we propose to use the sentiment analysis and social network analysis to mine the communities in the news video event.

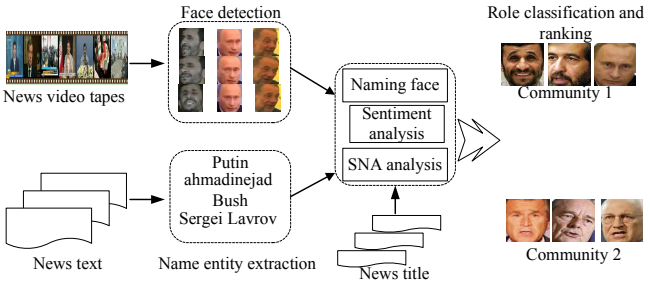

Fig. 2 Role classification and ranking framework.

The rest of the paper is organized as follows. Section 2 introduces the related work. Section 3 presents the news video naming faces approach. In section 4, we present our algorithm on news people/organization classification and community mining. Experimental results are reported and analyzed in section 5. Finally, we conclude the paper with future work in section 6.

2. RRELATED WORK

In this section we mainly review the related work on people relationship mining and social network analysis. Recently, social network analysis has become a hot research topic. Modeling the interaction between different entities by using complex network, the aim of social network analysis is to discover hidden structures or properties that cannot be directly perceived or measured by people. The idea of social network analysis has been used in many areas including internet structuring, human interactions [4], epidemiology, ecosystems [5], *etc*. More recently, some work has been contributed to multimedia analysis. A method [6] is proposed to analyze the relationship between the peoples in the news using their concurrent information. Another interesting work was proposed to analyze movie [7]. They treat a movie as a small society, which is constructed by roles and their interactions. Then, the relationships between roles are modeled by social network and they call it *RoleNet*. Based on the constructed *RoleNet*, several social network analysis algorithms were proposed to discover the hidden semantics, including leading roles detection, community identification, and movie story segmentation. The idea of this work is similar to our approach, however there are some differences. In their work, the relationships between different roles are analyzed by using concurrent information, the more concurrence the closer the two roles are. However, this assumption is not hold in the news domain, where not only roles in the same community but also the roles belong to different community concurrent in the same story. We will solve this problem by sentiment analysis. Another work similar to us is the *Renlifang* [8] proposed by *Microsoft Research Asia*. The engine can mine out the people who are close to the user's query. The relation intensity is reflected by the color of the name entities. Our approach is different from *Renlifang* in several aspects. *Renlifang* mainly focus on text analysis while our work analyzes news video by using multimodality information. *Renlifang* focus on mining the relationship between different entities through their concurrence information in the *WWW*, while our approach concentrates on analyzing the relationship of the entities according to a specific news event using sentiment analysis.

3. PEOPLE IDENTITY RECOGNITION

In order to mine the communities in the news event, we first identify the people appearing in the videos. Face recognition is the straight method to fulfill the task. However, the state-of-the-art face recognition method is based on statistical analysis and needs lot of data for training. The people are different from one topic to another and the pre-trained models for one topic cannot be transferred to another topic automatically. At the same time, labeling training samples is time consuming and labor intensive. Instead of using face recognition approach, we identify the people by associating their faces with names. Here, we adopt a simple but effective approach [9], which is scalable and needs no human labeled data for learning.

3.1. Face Detection and Representation

For naming face the first step is to obtain the faces and names. We adopted the method proposed by [10] to detect the faces, which can achieve good performance. In our approach, each news story is further divided into shots, and face detection is performed on each shot in every 5 frames interval. One important issue for naming face is how to represent each face. In our approach we adopt the local Gabor feature recognition approach in [11]. Before feature

extraction, the size of each face image is normalized to 128 by 160 pixels with the eye distance being 72 pixels. Totally 40 Gabor wavelets are used and the final feature dimension is 40960. Further these features are down-sampled and the dimension is reduced to 640.

3.2. Name Entity Related Positive Face Retrieval and Person Identification

The name entities are extracted from the news anchorperson speech text using the automatic name entity recognition tool [12]. After obtaining the entities, we use them as queries to retrieve name related faces using image search engine. The current image search engine is not perfect and the search result contains noise. On one hand, the wrong retrieved images may contain no faces. Therefore, we run face detection on each of the retrieved image, and the image with no face or more than one face will be discarded. On the other hand, the retrieved images may contain wrong faces. In order to select the right face images and eliminate the no-relevant ones, we proposed to use the manifold ranking approach [13] to re-rank the detected face, and select only the top ones as positive samples. The manifold ranking algorithm is initially proposed to rank data points along their underlying manifold, which is revealed by the relationship among the data points. Although such manifold ranking assumption may not hold in our situation, the way in which the relationship among the data points is investigated can be well applied to measure the relevance between the faces.

Assume there are *n* faces detected in the retrieval result. The face set can be represented as $\chi = \{x_1, ..., x_n\}$, where x_i represents the *i*th face in the initial ranking list. Let $d(x_i, x_j)$ represent the distance between the two faces x_i and x_j, and f denote a ranking function which assigns to each face x_i a ranking score f_i. We define a vector $y = [y_1, ..., y_n]^T$, where $y_i = 1$ if x_i is a labeled positive sample, otherwise 0. In our approach there are no labeled positive samples, and we assume the top *m* faces as the positive ones. Although the top *m* faces may contain noisy faces, the manifold ranking algorithm can overcome this noise by structure learning. The detail of the manifold ranking can be found in [13].

After the name related positive face samples are retrieved, the next step is to use these faces to match the faces detected in the videos. In our approach the face matching is under the story constraint, which means that the name related faces and the detected video faces are in the same news video story. We adopt the mean average distance to evaluate the dissimilarity of the faces, which is shown in equation (1).

$$d(F_i, F_j) = \frac{1}{M \times N} \sum_{f_m \in F_i} \sum_{f_n \in F_j} |f_m - f_n| \quad (1)$$

where $|\bullet|$ denotes the Euclidean distance; F_i and F_j are the two face sets detected from the video and web respectively, and f_m and f_n are the features of two faces belong to those two sets respectively. *M* and *N* are the face numbers.

4. NEWS VIDEO PEOPLE/ORGANIZATION CLASSIFICATION AND RANKING

In this section, we will mine the relationship between different organizations and classify the people/organizations into different communities. To represent the relationship between different organizations, two social networks based on the news stories are constructed. A graph is used to represent these relations, as shown in Fig.3.

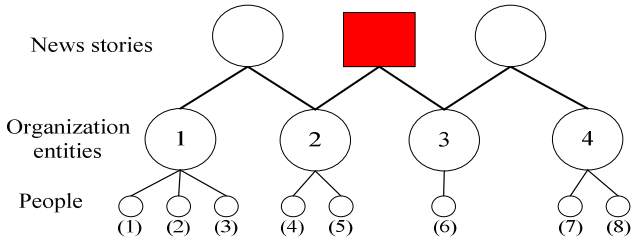

Fig. 3 A relation graph example.

The red square represents the relationship between these two entities in the news story is negative, while the white circles denote positive. The edge between a news story and an organization entity represents that the organization presents in this story. The edge between an organization entity and a person denotes the person appearing in the video belongs to that organization. We build two social networks $PG = (V, PE)$ and $NG = (V, NE)$, where V represents the set of organization entities, *PE* represents the set of positive edges and *NE* denotes the set of negative edges among the different country entities. We construct two matrixes $W_p = [a_{ij}^p]_{m \times m}$ and $W_n = [a_{ij}^n]_{m \times m}$ to record weight information between the different organizations, where a_{ij}^p denotes the number of positive stories the organization *i* and *j* occur together, and a_{ij}^n represents negative ones.

4.1. People Clustering According to Organization

Through observation, we find that the people appear in critical political news videos are usually the powerful people of a specific country or organization, such as the president, minister, news spokesman *etc*. The opinions expressed by those people not only represent the attitudes of their own, but also the opinions of the organizations which they belong to. For example, the opinion of *Bush* represents the attitude of the *U.S.A* and the opinion of *Tony Blair* always represents the attitude of the *Britain*. Each organization usually contains a few people (an example is shown in Fig. 4) and the opinions of them are usually uniform. Therefore, before classifying these people we can cluster the people belonging to the same organization.

It is straightforward to first recognize the relationship between the people and the organizations, and then cluster the people belonging to the same organization. Firstly we extract the candidate organization entities. As the people name entity, we extract the organization information from the news text. Assume the extracted organization entities are

$C_1, C_2,\ldots,C_N$, and the person names are $N_1, N_2,\ldots,N_m$. Their associations are also constrained by the story, which means that only the people names and the organization names appearing in the same story are analyzed. We infer the organization label of a specific person by calculating the Bayesian posterior probability in equation (2).

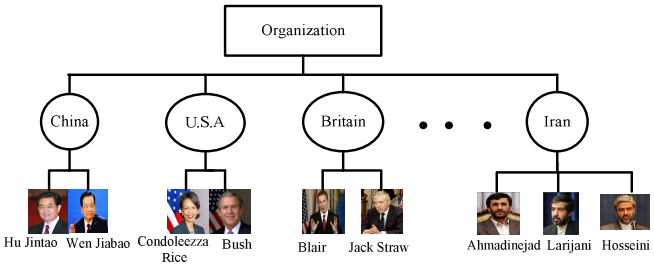

Fig. 4 Organization person example.

$$p(C_i / N_j) = \frac{p(N_j / C_i)p(C_i)}{\sum_{k=1}^{N} p(N_j / C_k)p(C_k)} = \frac{p(C_i, N_j)}{p(N_j)} \quad (2)$$

The final label of the person name *j* is determined by,

$$NC_j = \arg\max_{C_i}(p(C_i / N_j)) \quad (3)$$

The key point of equation (3) is how to calculate the probabilities. In our application, we have only a few news stories related to the same news event but do not have enough data to estimate these probabilities. However, we can estimate these probabilities by using extra information from the web. The assumption is that the more the person and the organization name concurrence in the web the closer relationship they are. We use the person and the organization name together as query to search in the *Google* and count the number of the returned websites. Let $Number_N$ represents the number of the returned websites for query *N*; $Number_{CN}$ represents the number of the returned websites for query *CN*. We assume the total number of website indexing by *Google* is *X*. Then equation (2) can be rewritten as:

$$p(C / N) = \frac{p(C,N)}{p(N)} = \frac{Number_{CN} / X}{Number_N / X} = \frac{Number_{CN}}{Number_N} \quad (4)$$

After these numbers and probabilities are obtained, we can obtain the organization labels of these people and fuse the people with the same label.

4.2. People/Organization Classification and Ranking

In the above subsection, we have classified the people into the organizations. If we can categorize those organizations with different opinions into different communities, the people can naturally be classified into different classes. In this subsection, we will first categorize these organizations into different classes by using the social network and sentiment analysis, and then classify those people according to organization labels. In order to calculate the values of a_{ij}^p and a_{ij}^n of the social networks defined above, we have to determine the sentiment orientation of each new story. In the news domain it is not easy to obtain the sentiment of the story. A news text example is shown below.

Title: *Bush Recognizes Kosovo Independence*

"*President Bush, touring Africa, said despite Russia's opposition, history will show the independence of Kosovo is "the correct move" and will bring peace to…….*"

The story is to show *Bush* back *Kosovo* independence. However, both positive word *correct* and negative word *opposition* appear in the text. Fortunately, we have news title to help us to mine the relationship between different entities. Though the length of the news title is short, it is the summary of the news and its sentiment is clear. From the news title of the above example, we can easily find that *Bush* support *Kosovo*. We first detect the name entities (person or organization name) in the title. Then, we map the person name into country name. If there is more than one entity, we determine the relationship between those entities by the sentiment orientation of the title. For sentiment analysis, we adopt the method proposed by Turney [14], which is simple but yet effective, to determine the sentiment value of each word and adopt the word with biggest positive or negative sentiment value as the sentiment indicator of the story. The sentiment value of a word is decided by,

$$ST(\text{word}) = \frac{1}{|Pwords|}\sum_{pword \in Pwords} PMI(word, pword) - \frac{1}{|Nwords|}\sum_{nword \in Nwords} PMI(word, nword) \quad (5)$$

where *Pwords/Nwords* represents a set of words with positive/negative sentiment, $|Pwords| / |Nwords|$ represents the number of positive/ negative words. The *Pwords* and *Nwords* are manually collected, which are

Pword={success, support, help, call, willing, welcome, consistent, advice, agree, hope, progress, peace, satisfaction, negotiations, accept, positive }

Nwords={criticize, warn, against, refute, illegal, no, accuse, threat, deny, sanction, revenge, violate, block, obstruction, refuse, reject, attack, cancel, abandon, negative }

When a new word comes, it is classified as having positive sentiment when *ST(word)* is positive and negative otherwise. The *PMI* in equation (5) is calculated as,

$$\text{PMI}(word_1, word_2) = \log_2\left\{\frac{p(word_1 \,\&\, word_2)}{p(word_1)p(word_2)}\right\} \quad (6)$$

where *p(word)* is the number of retrieved website by search engine. After the values of $W_p = [a_{ij}^p]_{m\times m}$ and $W_n = [a_{ij}^n]_{m\times m}$ are obtained, we mine the underline communities in the news based on the following assumptions: (1) friends of friends are friends; (2) friends of enemies are enemies; (3) enemies of friends are enemies; (4) enemies of enemies are friends. The influence of each organization is described by its centrality which is calculated as below,

$$c_j = (\sum_j a_{ij}^p + \sum_j a_{ij}^n) / (\sum_{ij} a_{ij}^p + \sum_{ij} a_{ij}^n) \quad (7)$$

The community mining approach is formulated as below.

Algorithm 1: The People/Organization Community Classification and Ranking Algorithm

1. Calculate the centrality value of each organization.
2. Sort the centrality values in descending order.
3. Label the organization with the biggest centrality value as 1 (assume the organization is *j*), and other roles 0.
4. If the label of organization *i* is 0, label it with -1 if $a_{ij}^p < a_{ij}^n$, otherwise 1.
5. For these organizations labeled in step 4, iterate until convergence.
6. Ranking the organizations in each community according to centrality values. The organizations with label 0 are deemed as other roles.
7. Classify the people into the communities according to their relationship to the organizations, and rank them according to their occurrence number in the event.

5. EXPERIMENTAL RESULT

5.1. Data Collection

In the experiment, six representative critical news topics are selected, which are: (1)"*Iran nuclear problem*, 2004", 189 stories; (2) "*North Korean nuclear six-party talks*", 139 stories; (3) "*Lebanon-Israel conflict*, 2006", 99 stories; (4) "*Russia and Georgia conflict*, 2008", 66 stories; (5) "*The Kosovo independence*, 2008", 27 stories; (6) "*US-Russian Anti-Ballistic missile issue*, 2008", 16 stories. They are typical critical international political events and have obviously underline communities. Those videos are collected from [15] together with the speech recognition text and the news video titles. The text is then translated into English for easy analysis [16].

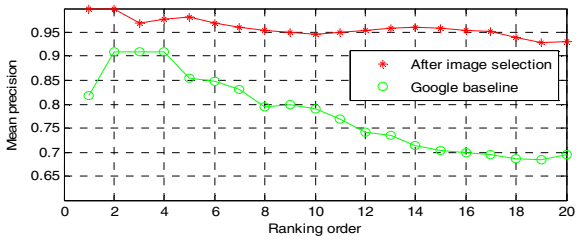

Fig. 5 Performance comparison before and after selection.

5.2. Face Image Retrieval and Naming Face

In the experiment, totally 1051 names and 1634 face clips are extracted. We use these names as queries to search in the *Google* image search engine, and explore the manifold ranking approach to select the right face images. Here *m* is set as 10 and *M* is set as 20. For some names, though some images are retrieved, there are no right related faces. We calculate the mean average distance between the selected top 20 faces and discard the candidates with low values. A threshold is defined according to experience, which is 0.55. The face precision comparison in the top 20 image before and after image selection for the final selected names is shown in Fig. 5. We can see the precision after image selection is much higher than of the baseline, and demonstrates the proposed selection approach is effective.

After the names related faces are extracted, we use them to match with the faces in each news video story and only the video faces and name related faces in the same story could be matched. A threshold is defined to determine if those faces are matched. The value of the threshold is selected as 0.52 according to experience. The experimental results of naming face on different topics are shown in Fig. 6. The naming face precision is calculated through (8).

$$Precison = \frac{|\text{correct named faces}|+|\text{correct discriminated face}|}{|\text{total faces}|} \times 100\% \quad (8)$$

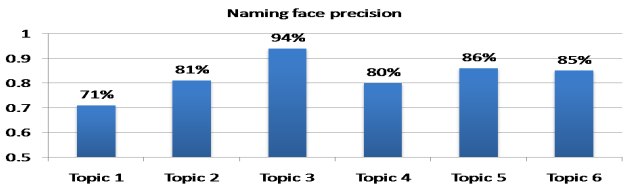

Fig. 6 The naming face result in each topic.

From Fig. 6 we can see that most of the experimental results are satisfactory. In the experiment, the main factor affecting our naming face performance is that some of the detected faces in the video are too small to match well with the retrieved faces.

5.3. People Related Organization Recognition

After the faces are named in the videos and the organization entities are extracted from the speech recognition text, the next step is to match the faces with the organizations. The matching results on the selected topics are shown in Fig. 7. The precision of the experiment is defined as,

$$Precision = \frac{\text{The number of right recognized person}}{\text{The number of total name}} \times 100\% \quad (9)$$

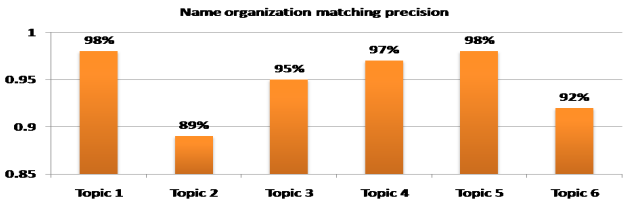

Fig. 7 The person organization matching result.

From Fig.7 we can see that the results are good and all of these precision are above 80%. The main factor that affects our matching may be that the *and* search operation is replaced by the *or* operation in the search engines.

5.4. Community Mining

In the experiment, we used the *PMI-IR* [14] method to evaluate each news titles and trained two thresholds to discard those news titles with non strong sentiment. Then by using the built social networks and the algorithm 1, the organizations are classified into different communities and ranked according to their influence, which is shown in Table 1. We manually checked the classification result and marked the wrong classified entities with Italic and red color. From the result we can see that the classification results fit well with the real situation.

Table 1. The organization classification result.

Topic	Community 1	Community 2	Others
1	Iran, IAEA, EU, Russia, Lebanon, Syria	United States, UN, France, Germany, Israel, Britain	China , Egypt
2	North Korea, South Korea, Russia, EU, IAEA	United States, Japan, UN, *China*	
3	Lebanon, UN, Syria, Iran, *China*	Israel, USA, France, *The Arab League*	Jordan
4	Russia, South Ossetia, Abkhazia, Germany, Venezuela	Georgia NATO, United States , EU	
5	Kosovo, NATO, the United States	Russia, Serbia	UN, EU
6	the United States, NATO, Poland	Russia	Czech, Bulgaria

After the communities in the organization level are mined out, the people can be classified into these communities accordingly. We can visualize those relationships by graph and the example of topic 6 is shown in Fig. 8. The red lines represent the negative news between the two entities and the green lines denote the positive ones. The size of the circle/face denotes the influence intensity of the organization/people in the event. From Fig. 8 we can see the relation between different organizations/people clearly and can browse our interesting news according to the people/organization/relationship accordingly.

6. CONCLUSION AND FUTURE WORK

In this paper, we propose a novel approach to mine the communities hidden in the news video event and classify different organizations/people into different communities. Firstly the famous elites appearing in the news video are detected. Then, the relationships between the organization entities are determined by mining the sentiment of the news title. At last the organizations and people are classified using social network analysis. The main contributions of the paper lie in that we propose a novel approach to present the news video event and propose to use the sentiment analysis to mine the relations between different organizations. The experimental results and the visualization example demonstrate the effectiveness of the proposed approach. In the future we may use the video itself to directly mine out the relationship between the different people.

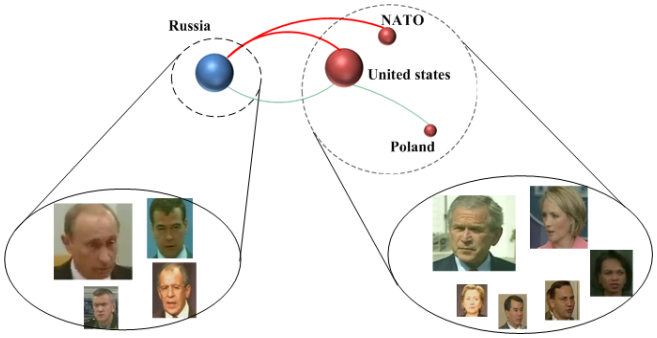

Fig. 8 The mining result visualization for topic 6.

7. REFERENCES

[1] T. S. Chua, S. F. Chang, L. Chaisorn and W. Hsu, "Story boundary detection in large broadcast news video archives – techniques, experience and trends," ACM Multimedia, 2004.

[2] V. S. Tseng, J. H. Su, J. H. Huang and C. J. Chen, "Integrated mining of visual features, speech features, and frequent patterns for semantic video annotation," IEEE Transaction on Multimedia, vol.10, no.2, pp.260-267, 2008.

[3] W. H. Hsu, L. S. Kennedy and S. F. Chang, "Video search reranking via information bottleneck principle," In Proc. ACM Multimedia, 2006.

[4] R. Guimera, L. Danon, A. Diaz-Guilera, F. Giralt, and A. Arenas, "Selfsimilar community structure in a network of human interactions," Phys. Rev., vol. 68, p. 065103(R), 2003.

[5] A. E. Krause, K.A. Frank, D. M. Mason, R. E. Ulanowicz, and W. W. Taylor, "Compartments revealed in food-web structure," Nature, vol. 426, pp.282–285, 2003.

[6] I. Ide, T. Kinoshita, H. Mo, N. Katayama, and S. Satoh, "Trackthem: exploring a large-scale news video archive by tracking human relations," AIRS, 2005

[7] C. Y. Weng, W. T. Chu and J. L. Wu, "RoleNet: movie analysis from the perspective of social networks," TMM, vol. 11, no.2, pp. 256-271, Feb. 2009.

[8] http://renlifang.msra.cn/

[9] C. X. Liu, S.Q. Jiang and Q.M. Huang, "Naming faces in broadcast news video by image Google," in Proc. ACM Multimedia, pp.717-720, 2008.

[10] J. Chen, X. Chen and W. Gao, "Expand training set for face detection by GA re-sampling," In Proc. IEEE int'l conf. on Automatic Face and Gesture Recognition, 2004.

[11] Y. Su, S. Shan, X. Chen and W. Gao, "Hierarchical ensemble of global and local classifiers for face recognition," In Proc. of IEEE Int'l Conf. on Computer Vision, pp.1-8, 2007.

[12] Alias-i. Lingpipe named entity tagger. In http://www.aliasi.com/lingpipe/.

[13] D. Zhou, O. Bousquet, T. N. Lal, J. Weston, B. Schoelkopf, "Learning with local and global consistency," *NIPS*, 2003.

[14] P. D. Turney etc, "Measuring praise and criticism: inference of semantic orientation from association," ACM Transaction on Information Systems, Vol.21, No.4, pp. 315-346, 2003.

[15] http://www.xinhuanet.com/

[16] http://www.google.cn/language_tools?hl=zh-CN