Cross Community News Event Summary Generation Based on Collaborative Ranking

Chunxi Liu$^{1,3}$, Weigang Zhang$^2$, Shuqiang Jiang$^3$, Qingming Huang$^{1,3}$

$^1$Graduate University of Chinese Academy of Sciences, Beijing, 100049, China
$^2$School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China
$^3$Key Lab of Intell. Info. Process., Inst. of Comput. Tech., CAS, Beijing, 100190, China

{exliu, wgzhang, qmhuang, sqjiang}@jdl.ac.cn

ABSTRACT
In order to make the users to access their interested news content conveniently, news analysis has been a hot research topic for a long time. However, most of the previous works only focus on news event detection, tracking, etc. Little attention has been paid to news report difference analysis and comparison. News is full of people’s cognition of the world. Because of the different background, ideologies, the cognition of the people is different from one to another. Then, the angle of the news report is different accordingly. In this paper we propose a novel scheme to summarize and compare news report from different communities by using news pictures. Two challenges are addressed: similar news reports summary generation and different news reports summary generation. Firstly, news pictures from different sources are downloaded. Then, the bag-of-visual word features are extracted from these pictures to represent their content. After that, based on the similarity and dissimilarity of the pictures from different sources, collaborative ranking is adopted to rank the images to obtain the news similar and difference summaries. Experimental results on the selected news topics are promising and demonstrate that the proposed approach is effective.

Categories and Subject Descriptors
H.3.3 [Information Storage and Retrieval]: Information search and retrieval-search process.

General Terms

Keywords
News summarization, co-ranking, event analysis.

1. INTRODUCTION
News plays an important role in our daily life. With the fast development of the modern technologies, such as computer technologies, television, internet etc, we can access the event happened in the world through news broadcasting immediately. Everyday large amount of news data, including text, audio, image and video, are generated, edited, and spread. Among these data, news pictures are often used as summary to show the events with the belief that “One picture is worth ten thousand words”. Therefore, it is of significant importance to analyze the news event from these news pictures.

The goal of the news reports is to describe the happened event objectively. However, the news data are generated by people with a specific background. Therefore, the reports are inevitably influenced by the author’s subjective cognition of the world. And for some events the reports are heavily influenced by the author’s prejudice. One obvious phenomenon caused by this bias is that for the same news event the reports of different community would be rather different. Such as the reports about the event “Kosovo independence”, which are shown in Figure 1. From these pictures we can see that the English reports focus on the people’s celebrations, while the Russian reports focus on the people and meeting. This may due to that the western countries back Kosovo independence while Russia opposes it. Because of the different viewpoints, the western and the Russia would like to select different pictures to represent their understanding about the event.

![Figure 1. Different source news picture comparison.](image_url)

While news reports difference widely exists in the real world, little research attention has been paid to it. Existing research on news analysis mainly concentrates on topic detection [1] and tracking [4], structure analysis [2], semantic concept detection, annotation [6], search [7], and monitoring [4]. Through structure analysis, the individual news stories are segmented and news video can be accessed by story level indexing. After concept detection and annotation, the semantics in the stories and videos can be understood. News video search can retrieve the user concerned stories. All of these research topics address the objective aspect of the news report. In this paper, we propose a novel approach to summarize and compare the reports from different sources using news pictures. This work is motivated by our previous work [8] and the work proposed by Wan [9]. In Wan’s work, they propose to summarize the differences in multilingual news text reports. Constrained co-ranking method is proposed to select common and different sentence to generate the document summary. Different from the work in [9] in this paper we focus on summarizing and comparing news using news pictures which is more vivid. In [10], the visual patterns from multi-engines are used to refine the search result. The idea of their
work is similar ours. While they utilize visual patterns from multi-sources to get a single ranking list, we propose to use the visual patterns from multi-sources to obtain both the similar and difference ranking lists. Another related work is proposed by Lin [3] in which they address the different ideological perspective problem of the news videos based on the emphatic patterns. The assumption is that news broadcasts holding contrasting ideological beliefs appear to emphasize different subsets of visual concepts. While we consent on the point that different news video channels usually focus on different aspect of happened news event, we cannot agree that they do always hold contrasting ideologies. The proposed news picture summarizing and comparing framework is shown in Figure 2.

Figure 2. The proposed news picture summarizing and comparison framework.

In this paper we will take news report pictures from the Chinese and English community as an example to show how to generate the similar and the difference news summaries. For a specific event such as “London riot” the news pictures from different communities are collected using the image search engine. Then, the bag-of-visual word image features are extracted from the images to represent their contents. After that, collaborative ranking is utilized to obtain the similar and the difference news picture summaries based on the relationship of the two news picture sets. Experimental results in the following sections will demonstrate the effectiveness of the proposed scheme.

The rest of the paper is organized as follows. Section 2 presents the news picture feature extraction and the similarity calculation. In section 3, the similar/difference news picture summary generation algorithm is described in detail. In section 4, the experimental results are reported and analyzed. Finally, we conclude the paper in section 6.

2. FEATURE EXTRACTION AND SIMILARITY CALCULATION

How to extract effective visual feature from the image has been a hot research topic for a long time. However due to the complexity of the problem until now there is no unified solution. Recently, a popular approach is proposed to resemble the image with the text [11], where it tries to extract the bag-of-visual word feature to represent the image content. It has been proved to be effective for image classification [12] and retrieval. In this paper the bag-of-visual word approach is adopted to conduct the clustering. The hierarchical K-means is of high efficiency and can organize the generated visual words in the vocabulary tree. With the hierarchical structure, finding the closest visual word for a local feature descriptor can be performed very efficiently. More details about the vocabulary tree and its applications can be found in [14].

Assuming that the image sets are represented as $\text{Image}=(I_1, I_2, ..., I_k)$, $k=(k_1, k_2, ..., k_k)$ is the dictionary of all visual words appearing in the whole image pool. After the bag-of-visual words are extracted, we adopt the vector model to describe the images, where the feature of an image is defined as:

$$f = (k, w)$$  

where $w=(w_1, w_2, ..., w_n)$ is a set of corresponding weights, $n$ is the number of unique visual words in the image dataset. We employ the term frequency ($tf$) to calculate the importance of a visual word in an image. $tf$ is the raw frequency of a given term inside an image. In this paper the cosine distance is adopted to measure the similarity between two images. Assuming the bag of visual word features of the two images are $dx$ and $dy$. The cosine distance between $dx$ and $dy$ is shown in equation (2), where $w(d_s)$ denotes the weight of $d_s$.

$$\text{Similarity}(d_s, d_o) = \frac{w(d_s) \cdot w(d_o)}{|w(d_s)| \cdot |w(d_o)|}$$  

3. NEWS SUMMARY GENERATION

In this section, the pictures in the two image sets are simultaneously ranked in a unified graph-based algorithm to obtain the news summaries. Each news picture is assigned a similar/dissimilar score to indicate how much the picture contains important and similar/difference information. The similar/dissimilar score of the Chinese/English news pictures relies not only on the Chinese/English news pictures but also on the English/Chinese news pictures. In particular, the proposed news similar/dissimilar summary generation method is based on the following four assumptions:

Assumption (1): The similar score of a Chinese/English picture should be high if it is heavily correlated with other Chinese/English pictures with high similar scores in the Chinese/English news picture set.

Assumption (2): The similar score of a Chinese/English picture should be high if the picture is heavily correlated to the English/Chinese pictures with high similar scores in the English/Chinese news picture set.

Assumption (3): The difference score of a Chinese/English picture should be high if the picture is heavily correlated with other Chinese/English pictures with high difference scores in the Chinese/English news picture set.

Assumption (4): The difference score of a Chinese/English picture should be high if the picture is heavily un-related with the English/Chinese pictures with high difference scores in the English/Chinese news picture set.

Let $P_m$ denote the English news picture set, $P_e$ denote the Chinese picture set. Let $G=(V_e, V_m, E_e, E_m)$ be the undirected graph for the pictures in the two image set. $V_e = |P_e|$, $|1 \leq c \leq m|$ is the set of English picture set. $V_m = |P_e|$, $|1 \leq c \leq n|$ is the set of Chinese picture set. $m$ and $n$ are the picture numbers in the two image sets, respectively. $E_e$ is the edge set to reflect the relationship between
the English news pictures, \( E \) is the edge set to reflect the relationship between the Chinese news pictures, and \( F \) is the edge set to reflect the relationship between the English news pictures and the Chinese news pictures. Based on the graph representation, the following three matrices are computed.

\[ M^+ = (M^+) \] represents the similarity between the English news pictures. \( M^- = (M^-) \) represents the similarity between the Chinese news pictures. \( M^0 = (M^0) \) represents the similarity between the English news pictures and the Chinese news pictures.

All the similarities in the above matrices are computed as in equation (2). Two column vectors \( u \) and \( v \) are used to denote the ranking scores of the Chinese news pictures and the English news pictures, respectively. The assumption (1) and (2) can be rendered as follows:

\[
u(s^c) = \alpha \sum M^+_c u(s^c) + \beta \sum M^-_c v(s^c)
\]
\[
u(s^e) = \alpha \sum M^+_e v(s^e) + \beta \sum M^-_e u(s^e)
\]

By fusing the above functions we can obtain:

\[
u(s^c) = \alpha \sum M^+_c u(s^c) + \beta \sum M^-_c v(s^c)
\]
\[
u(s^e) = \alpha \sum M^+_e v(s^e) + \beta \sum M^-_e u(s^e)
\]

where \( \alpha \) and \( \beta \) are the two tradeoff parameters and \( \alpha + \beta = 1 \). The two equations are computed iteratively until convergence. In order to guarantee the convergence of the iterative form, \( u \) and \( v \) are normalized in each iteration. The news picture similar summary generation algorithm is summarized as bellow.

### Similar news summary generation

1. Calculate the similarity between the English news pictures to obtain \( M^{en} \).
2. Calculate the similarity between the Chinese news pictures to obtain \( M^{cn} \).
3. Calculate the similarity between the Chinese news images and the Chinese news images to obtain \( M^{cen} \).
4. Initialize \( u^t = u(u^0, ..., u^0) \) and \( v^t = v(v^0, ..., v^0) \) randomly such that \( \sum u^t = 1 \), \( \sum v^t = 1 \) and \( u^t > 0 \), \( v^t > 0 \).
5. Set \( t = 1 \).
6. Compute the news picture similar summary scores

\[
u'(s^c) = \alpha \sum M^+_c u'(s^c) + \beta \sum M^-_c v'(s^c)
\]
\[
u'(s^e) = \beta \sum M^+_e u'(s^e) + \alpha \sum M^-_e v'(s^e)
\]

\[
u^t \leftarrow \frac{\nu'}{||\nu'||}, \quad v^t \leftarrow \frac{\nu'}{||\nu'||}.
\]

7. \( t \leftarrow t + 1 \).
8. Repeat step 6 to 8 till convergence (i.e., \( \nu^t = \nu^{t+1}, v^t = v^{t+1} \)).
9. Select the top ranked news pictures as the similar summary for each news picture set.

### 4. EXPERIMENTAL RESULT

In this section, extensive experiments are conducted, including data collection, parameter setting, similar news picture summary generation and difference news picture summary generation.

#### 4.1 Data Collection and Parameter Setting


These topics are typically international political events which attract much social attention. In order to collect the news event related pictures the Google image search engine is adopted. We select Google image [5] for the reasons that Google image is the most popular image search engine cross many language communities. For Google image, after retrieval only the top1000 images could be assessed, which contain all kinds of image. Those images whose resolutions are smaller than 125x125 will be discarded. The parameters \( \alpha \) and \( \beta \) in the ranking algorithms are both set as 0.5 according to experience.

#### 4.2 Similar News Summary Generation

In this subsection, the similar news picture summaries for the selected eight topics are generated according to the algorithm described in section 3. The summary generation results are shown in Figure 3. From these summaries we can see that the summaries for the same event roughly reflect the same meanings. For example, the summaries for Iran both cover the Uranium enrichment plant. For topic Kosovo and Libyan the summaries both cover protesting and celebration, and for Mekong tragedy the summaries both reveal the people cleaning the bodies beside the river. For Thailand the summaries both mainly reflect the scenes of the flood in the city. It should be noted that the Chinese
summary for Syrian riot consist of five near duplicate pictures which are different from the pictures in the English summary. In our view it may be caused by the fact that when computing the ranking scores, the near duplicate images reinforce each other and obtain high ranking scores when the algorithm converges.

4.3 Difference News Summary Generation
Similar to subsection 4.2, the difference news picture summaries for the selected eight topics are shown in Figure 4. From these summaries we can see that different from the similar summaries, the difference summaries for the same event in Figure 4 reflect different meanings. The difference focus comparison between these two summaries is shown in Table 1. From the results we can see that these summaries really focus on different part of the event. For example for the topic "European crisis" the English focus on the stock market while the Chinese focus on other parts of the event. For the topic “Mekong tragedy” the English focus on the Mekong river while the Chinese focus on the tragedy. For the topic “Flooding in Thailand” the English focus on the floods while the Chinese focus on the hard disk production decrease and the stock market.

![Figure 4. The difference news summaries for the 8 topics.](image)

![Table 1: The focus comparison of the two summaries.](table)

<table>
<thead>
<tr>
<th>Topics</th>
<th>English summary</th>
<th>Chinese summary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Iran</td>
<td>news writing</td>
<td>locations and meetings</td>
</tr>
<tr>
<td>Kosovo</td>
<td>flags</td>
<td>maps</td>
</tr>
<tr>
<td>Syrian riot</td>
<td>riot and fires</td>
<td>maps and others</td>
</tr>
<tr>
<td>London riot</td>
<td>police and fires</td>
<td>other figures</td>
</tr>
<tr>
<td>European crisis</td>
<td>stock market</td>
<td>other parts of the event</td>
</tr>
<tr>
<td>Libyan</td>
<td>protesting</td>
<td>Libyan maps</td>
</tr>
<tr>
<td>Mekong tragedy</td>
<td>river</td>
<td>Chinese concern about the event</td>
</tr>
<tr>
<td>Floods in Thailand</td>
<td>floods</td>
<td>hard disk production decrease, stock market</td>
</tr>
</tbody>
</table>

5. CONCLUSION AND DISCUSSION
The major contribution of the paper is that we propose a novel approach to generate similar and difference news picture summaries for the news data from different sources. Although the result is encouraging, there are some problems. Firstly, the data collected by Google image contain noise, which influences the performance of the summary generation. In the future we will try to collect news image data from other ways to ensure the quality of the images. Secondly, the bag-of-visual word feature cannot capture the whole semantic of the picture. In the future we will explore more effective feature to represent the image content. Thirdly, the proposed summary generation approach only considers the relationship between the images and neglects some problems. For example in summary the near duplicate images should be removed to contain more information. Finally, in the future we will try to add more quantitative and qualitative experiment to justify the effectiveness of the proposed summary generation scheme.

6. ACKNOWLEDGMENTS
This work was supported in part by National Basic Research Program of China (973 Program): 2012CB316400, in part by National Natural Science Foundation of China: 61025011, 60833006 and 61071018.

7. REFERENCES