MOVIE ANALYSIS BASED ON ROLES' SOCIAL NETWORK

Chung-Yi Weng, Wei-Ta Chu, and Ja-Ling Wu

Department of Computer Science and Information Engineering
Graduate Institute of Networking and Multimedia
{chunye,wtchu,wjl}@cmlab.csie.ntu.edu.tw

ABSTRACT

Roles in a movie form a small society and their interrelationship provides clues for movie understanding. Based on this observation, we present a new viewpoint to perform semantic movie analysis. Through checking the co-occurrence of roles in different scenes, we construct a roles' social network to describe their relationships. We introduce the concept of social network analysis to elaborately identify leading roles and the hidden communities. With the results of community identification, we perform storyline detection that facilitates more flexible movie browsing and higher-level movie analysis. The experimental results show that the proposed community identification method is accurate and is robust to errors.

1. INTRODUCTION

Due to vigorous development of movie industries, more than 4500 movies are produced every year. With advance of digital technologies, movies are produced or disseminated digitally, and seeing movies has been one of the most popular entertainments. Therefore, users pose urgent needs on organization and indexing techniques for efficient access of massive movie data.

Many studies have been proposed to analyze movies based on audiovisual features. They can be roughly categorized into the following fields: genre classification, story segmentation, and video abstraction. Rasheed et. al. [1] exploit color, motion, and shot information to classify movies into comedies, action, dramas, or horror films. Based on similar audiovisual features, Adams et. al. [2] evaluate video tempo to characterize movies. Also, visual features are widely applied to describe logical story units [3], which are series of shots that convey similar semantics. Various video abstraction techniques [4] have been proposed to represent movie content in a compact manner.

Some researches tackle with movie analysis from different perspectives. The so-called "affective content analysis" investigates human's perception drawn by audiovisual stimuli [5][6]. However, the reported methods still rely on modeling low-level features and don't approach to "understand" movie content.

In this work, we propose a novel way to analyze movie content. We understand the stories conveyed by a movie because we know the mutual relations between roles. How the roles interact or conflict leads the story. Therefore, to simulate the thinking progress of human in seeing a movie, we advocate that we should analyze roles' social relationships.

The roles in a movie form a small society, which may be further divided into several communities. For example, in many action movies, there are often a community representing justice and another community representing evil. Identifying roles' relationship and communities would provide significant clues in movie analysis and let computers understand more about the hidden semantics.

We introduce techniques that are originally developed for social network [9] to movie analysis. We construct a roles' social network and identify the hidden communities. Based on this information, how the director narrates a story can be derived in a more reasonable way. The proposed method appropriately approaches how humans see movies and more elaborately analyze movies beyond conventional feature-based approaches.

Contributions of this work are summarized as follows:

- We propose a new way to analyze movie content, through examining the social relationships between roles.
- We introduce the concept of social network analysis to multimedia content analysis.
- We demonstrate the effectiveness of using the roles' social network in various applications.

The rest of this paper is organized as follows. Section 2 describes the proposed framework. Section 3 describes construction details of roles' social network and community identification. Based on this network, we propose an application called storyline detection in Section 4. Experimental results are shown in Section 5. Section 6 gives some discussions on the proposed approach and Section 7 concludes this paper.

2. THE PROPOSED FRAMEWORK

Based on the idea of utilizing roles' mutual relationship, we propose a framework to analyze movies, as shown in Figure 1. At the scene labeling stage, scene boundaries are determined, and the roles show up in each scene are identified. Based on labeled scenes, roles' relationships are transformed as a "roles' social network." The proposed movie analysis then acts on this network, including determining the leading roles and community identification. Once we find roles' relationships and corresponding community, we can apply them to develop high-level applications, such as storyline detection or highlighted scenes analysis.

Robust scene boundary detection and role recognition in movies are still very challenging problems. Although many methods have been proposed, none of them is perfect in general. In order to particularly evaluate the proposed idea, we manually perform scene detection and face recognition as the foundations for the proposed processes.

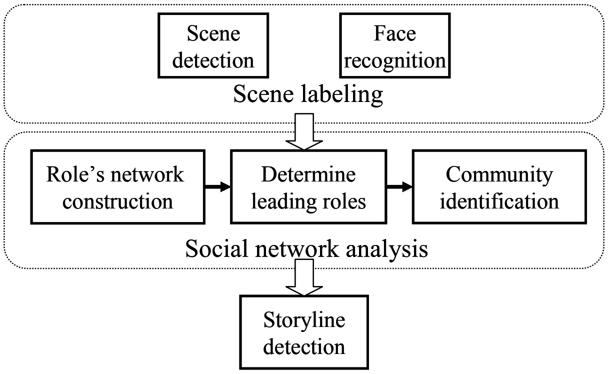

Figure 1. Structure of the proposed framework.

3. SOCIAL NETWORK ANALYSIS

3.1. Roles' Social Network Construction

To detect the relationship between roles, we construct a roles' social network based on labeled scenes. A bipartite graph is firstly used to represent labeled scenes, as shown in Figure 2. The square nodes represent scenes, while the circular nodes represent the roles in this movie. An edge between a scene node and a role node denotes that the role presents in this scene.

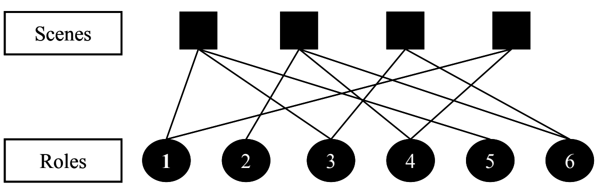

Figure 2. An example of bipartite graph.

A roles' social network $G = (V, E)$, where V denotes the set of roles presenting in the movie and E denotes the set of edges among the roles, is constructed based on the bipartite graph. If both the ith role and the jth role connect to the same scene in the bipartite graph, an edge e_{ij} between v_i and v_j is added to E. Moreover, we construct a weight matrix W to record weight information, which stands for the number of scenes where two roles co-occur. Note that an element w_{ij} is equal to w_{ji}. Weight information represents the closeness between two roles.

Figure 3 is a real case of the network constructed for the movie called "You've Got Mail". The nodes represent roles, and thickness of edges represents the corresponding weight information. Basically, "You've Got Mail" is a typical "bilateral movie", in which there are two leading roles (the hero and the heroine), and roles surround them form two communities. Table 1 shows the overview of roles' relationships in this movie. However, these characteristics cannot be directly figured out from Figure 3, which shows intricate edges between nodes. Some manipulations must be made to facilitate deeper investigation. Fortunately, many social network analysis techniques are widely studied in sociology. We apply these techniques to roles' social network and detect the leading roles and hidden communities.

We believe that analyzing bilateral movies is a good start in movie analysis. The reason is that it's very common in popular movies. For example, there are justice and evil communities in most action movies. As for romance movies, it is very common to find two communities that belong to the hero and the heroine, respectively.

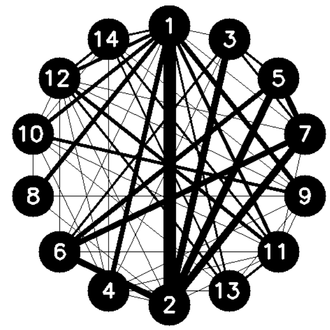

Figure 3. Roles' social network of "You've Got Mail".

Table 1. Roles in "You've Got Mail".

Node	**Meaning of roles**
1	The hero (Tom Hanks)
2	The heroine (Meg Ryan)
3, 5, 6, 7	The heroine's friends and colleagues
4, 8, 9, 10, 11, 12, 13, 14	The hero's friends, relatives, and colleagues.

3.2. Determine Leading Roles

Leading roles are the persons who have significant impacts in movies. In social network analysis [9], evaluating impact of each individual is one of the earliest issues to be pursued. It is known as the *centrality problem*. In general, centrality of a node can be evaluated as the number of connected edges. However, this method ignores much useful information embedded in the weights of edges. Therefore, we modify the measurement method and calculate the centrality c_i of a node i as

$$c_i = \sum_j w_{ij}. \tag{1}$$

Because we mainly conduct bilateral movies, we choose the nodes with the first two largest centrality values as the leading roles. They are briefly denoted as v_p and v_q in this paper.

3.3. Community Identification

To accomplish more advanced movie analysis, one of the effective ways is to further understand the relationships between roles. In social network analysis, discovering roles' relationships can be achieved by community identification. Communities are groups of nodes within which the connections are dense but between which they are sparser.

According to the characteristics of bilateral movies, there are two major communities led by two leading roles. Determining these two communities can be viewed as a binary labeling problem.

Given a roles' social network $G = (V, E)$ and corresponding weight information W, find a labeling solution $\boldsymbol{\Delta}^*$ as follows:

$$\Delta^* = \arg\min_{\Delta} C(\Delta) \quad \text{subject to } \delta_p = 0 \text{ and } \delta_q = 1, \tag{2}$$

$$C(\Delta) = \sum_{(i,j)\in E} \left|\delta_i - \delta_j\right| w_{ij}, \tag{3}$$

$$\Delta = \{\delta_i,\ i = 1, 2, \dots, n\}, \tag{4}$$

$$\begin{cases} \delta_i = 0 & \text{if } v_i \text{ is assigned to the community led by } v_p \\ \delta_i = 1 & \text{if } v_i \text{ is assigned to the community led by } v_q \end{cases}$$

where n is the number of roles, and $C(\mathbf{\Delta})$ is the closeness between two communities, which is calculated by summing the weights between roles in two different communities. The first leading role v_p is labeled as 0 (δ_p=0), and the second leading role v_q is labeled as 1 (δ_q=1). For brief description, we use $\mathbf{T}_p$ to represent the roles in the community led by v_p, and use $\mathbf{T}_q$ to represent the roles in the community led by v_q.

This problem can be solved through finding the minimum cut between two leading roles. Therefore, we adopt the maximum-flow-minimum-cut algorithm [8] in this work.

Figure 4 shows the community identification result of the network in Figure 3. Node 1 and node 2 are correctly detected as the leading roles. The roles identified as the members of $\mathbf{T}_p$ are visualized as circles, and the roles identified as the members of $\mathbf{T}_q$ are visualized as squares. These results exactly match the real cases listed in Table 1.

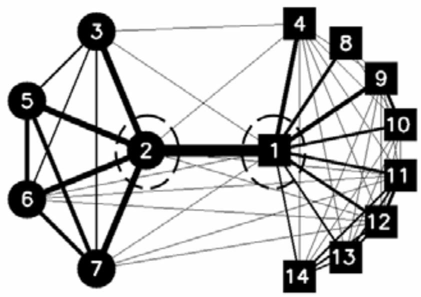

Figure 4. Community identification result of Figure 3.

4. STORYLINE DETECTION

In a movie, different storylines often interleave throughout the whole film. A storyline refers to the activities or events derived by roles with common attributes. For example, in action movies, directors often describe justice and evil communities alternately, thus constructing two different storylines.

Figure 5 shows an example of the romance movie "You've Got Mail". Because it's a bilateral movie, there are clearly two storylines that respectively describe the social contacts within communities. Therefore, we can utilize the results of community identification to easily detect storylines, which is hardly done in conventional feature-based movie analysis approaches.

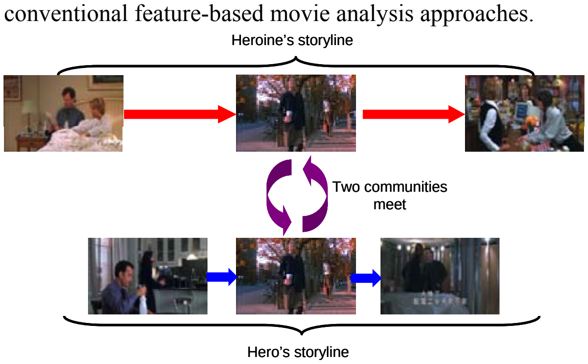

Figure 5. Parts of storylines in the movie "You've Got Mail".

To detect storylines, we classify scenes into one of the three sets S_1, S_2, and S_3. They are initialized as empty set Ø.

1) If all roles in the scene s_i belong to $\mathbf{T}_p$, then $S_1 = S_1 \cup \{s_i\}$.
2) If all roles in the scene s_i belong to $\mathbf{T}_q$, then $S_2 = S_2 \cup \{s_i\}$.
3) If parts of roles in s_i belong to $\mathbf{T}_p$, while others belong to $\mathbf{T}_q$, then $S_3 = S_3 \cup \{s_i\}$.

When all scenes are classified, we simply concatenate the scenes in S_1 and S_3, and sort them according to temporal order to construct the first storyline. The second storyline is composed of scenes in S_2 and S_3. Detecting storylines facilitates viewers to watch movies from different viewpoints and pave a new way for movie analysis.

5. EXPERIMENTAL RESULTS

We evaluate the proposed community identification method and its robustness to labeling errors, and show the precision of the storyline detection method. Three Hollywood movies are used, as shown in Table 2.

Table 2. Evaluation data.

Movie Title	Duration	Genre
You've Got Mail	119 min	Comedy/Romance
Mr. & Mrs. Smith	120 min	Action/Romance/Comedy
A Lot Like Love	107 min	Comedy/Romance

5.1 Community Identification

To evaluate the precision of community identification, we manually label which role should belong to which community. The identification precision is listed in Table 3. We can see that almost perfect performance can be achieved. This reflects that although movies seem to be very complex in aural or visual characteristics, the ways that directors tell a story often appeal to roles' social relationships. This kind of relationship can be clearly identified by checking how the roles show out in scenes.

Table 3. Precision of community identification.

Movie Title	# roles correctly labeled / # roles labeled	Precision
You've Got Mail	14 / 14	100%
Mr. & Mrs. Smith	12 / 13	92.3%
A Lot Like Love	14 / 14	100%

5.2 Robustness of Community Identification

Discovering social relationships between roles relies on correct scene labeling, which consists of scene boundary detection and face recognition. However, face recognition techniques are still not perfect up to now. Therefore, we evaluate how the face recognition errors affect the proposed community identification results.

To simulate the behavior of a non-perfect face recognition module, we artificially attack scene labeling data to simulate recognition errors. Errors are classified as:

1) Insertion: non-human objects are mis-recognized as human.
2) Replacement: a person is mis-recognized as another person.
3) Deletion: human faces are misdetected.

We randomly choose a scene and randomly attack it by one kind of these errors. We evaluate robustness of the community identification process by applying different strength of attacks. The strength of an attack is defined as:

$$e(\lambda) = \left(\sum_{i=1}^{k} N(s_i) \right) \times \lambda , \tag{5}$$

where k is the number of scenes in the movie, $N(s_i)$ is the number of roles in the scene s_i. The first term of the right part of eq. (5) is therefore denotes the total number of appearance by all roles. λ is the ratio of error-infected appearances to total appearances. Larger value λ is, stronger errors are applied. To comprehensively evaluate robustness, we check the precision of community identification under different attack strengths ($0 \leq \lambda \leq 1$). For a specific λ, we apply it α times and calculate the average precision as

$$\frac{1}{\alpha}\sum_{i=1}^{\alpha} p(e_i), \tag{6}$$

where $p(e_i)$ denotes the precision of the community identification result in the ith attack. α is the number of attacks, which is set as 100 in this evaluation.

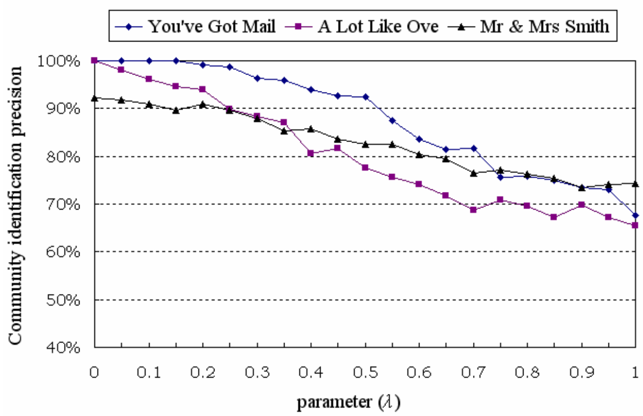

Figure 6. Robustness of community identification.

Figure 6 shows the evaluation results. We can see that different movies have different robustness, while the community identification is robust in general. Even the value of λ is up to 0.25, the precision of community identification is higher than 88%. The reason for this is that directors often clearly present important communities. A few face recognition errors would not significantly affect the community result.

5.3 Storyline Detection

The precision of storyline detection is based on the precision of scene classification described in Section 4. Therefore, we compute the precision of scene classification in this evaluation. The ground truth of scene classification is manually defined, and the experimental result is shown in Table 4. The numbers n/m in each cell denote that n scenes are correctly classified in m classified scenes. On the basis of precise community identification results, almost perfect scene classification performance is achieved.

Table 4. Performance of storyline detection.

Movie title	S_1	S_2	S_3
You've Got Mail	19/19	22/22	17/17
Mr. & Mrs. Smith	19/19	17/17	36/36
A Lot Like Love	15/15	19/19	36/36

6. DISCUSSION

We have shown that the proposed method operates quite well for extracting more semantic information. Moreover, we believe that exploiting roles' social relationships has much potential in movie analysis. Some exciting extensions of this idea are addresses as follows.

1) We can discover hidden semantics of each scene by observing the community structures. For example, some scenes represent the interaction between two communities, and some scenes enhance the relationships between roles in the same community. This kind of subtle information facilitates more elegant highlight and summary generation.
2) Through analyzing the interactions between different storylines, automatic movie recreation becomes feasible.
3) It's important to point out that the proposed method and existing ones are not irreconcilable. They can be combined to achieve finer movie understanding.

7. CONCULSION

This paper presents a novel way to perform movie analysis. We construct a roles' social network and identify the embedded community structure in a movie. On the basis of community information, we elaborately detect storylines, which is hardly achieved by existing methods. The experimental results show that the effectiveness of the proposed method and its robustness to face recognition errors.

In the future, we would like to combine conventional methods to develop more interesting applications. Another direction is to investigate more generalized algorithms which can be applied to various kinds of movies.

8. REFERENCES

[1] Z. Rasheed, Y. Sheikh, and M. Shah, "On the use of computable features for film classification," IEEE Trans. on Circuits and Systems for Video Technology, vol. 15, no. 1, pp. 52-64, 2005.

[2] B. Adams, C. Dorai, and S. Venkatesh, "Toward automatic extraction and expression of expressive elements from motion pictures: tempo," IEEE Trans. on Multimedia, vol. 4, no. 4, pp. 472-481, 2002.

[3] J. Vendrig and M. Worring, "Systematic evaluation of logical story unit segmentation," IEEE Trans. on Multimedia, vol. 4, no. 4, pp. 492-499, 2002.

[4] Y. Li, S.-H. Lee, C.-H. Yeh, and C.-C. J. Kuo, "Techniques for movie content analysis and skimming: tutorial and overview on video abstraction techniques," IEEE Signal Processing Magazine, vol. 23, no. 2, pp. 79-89, 2006.

[5] H.L. Wang and L.-F. Cheong, "Affective understanding in film," IEEE Trans. on Circuits and Systems for Video Technology, vol. 16, no. 6, pp. 689-704, 2006.

[6] A. Hanjalic and L.-Q. Xu, "Affective video content representation and modeling," IEEE Trans. on Multimedia, vol 7, no. 1, pp. 143-154, 2005.

[7] B. Jung, T. Kwak, J. Song, and Y. Lee, "Narrative abstraction model for story-oriented video," Proc. of ACM Multimedia Conference, pp. 828-835, 2004.

[8] Y. Boykov and V. Kolmogorov, "An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision," IEEE Trans. on Pattern Analysis and Machine Intelligence, pp. 1124~1137, 2004.

[9] J. Scott, "Social network analysis: a handbook," Newbury Park, 1991.