



Polibits

ISSN: 1870-9044

polibits@nlp.cic.ipn.mx

Instituto Politécnico Nacional

México

Kundu, Amitava; Das, Dipankar; Bandyopadhyay, Sivaji
Scene Boundary Detection from Movie Dialogue: A Genetic Algorithm Approach
Polibits, vol. 47, 2013, pp. 55-60
Instituto Politécnico Nacional
Distrito Federal, México

Available in: <http://www.redalyc.org/articulo.oa?id=402640461007>

- How to cite
- Complete issue
- More information about this article
- Journal's homepage in redalyc.org

redalyc.org

Scientific Information System

Network of Scientific Journals from Latin America, the Caribbean, Spain and Portugal

Non-profit academic project, developed under the open access initiative

Scene Boundary Detection from Movie Dialogue: A Genetic Algorithm Approach

Amitava Kundu, Dipankar Das, and Sivaji Bandyopadhyay

Abstract—Movie scripts are a rich textual resource that can be tapped for movie content analysis. This article describes a mechanism for fragmenting a sequence of movie script dialogue into scene-wise groups. In other words, it attempts to locate scene transitions using information acquired from a sequence of dialogue units. We collect movie scripts from a web archive. Thereafter, we preprocess them to develop a resource of dialogues. We feed the dialogue sequence from a script to a Genetic Algorithm (GA) framework. The system fragments the sequence into adjacent groups of dialogue units or output ‘scenes’. We use SentiWordnet scores and Wordnet distance for dialogue units to optimize this grouping so that adjacent scenes are semantically most dissimilar. Then we compare the resulting fragmented dialogue sequence with the original scene-wise alignment of dialogue in the script.

Index Terms—Dialogue, genetic algorithm, movie script, scene.

I. INTRODUCTION

WITH loads of digital movies being consumed online, movie search has become an important concern in the information retrieval domain. A lot of studies have reported extraction of semantically meaningful information from movie data [1–4]. Such information is useful in browsing, searching, and genre classification. Feature-length movies usually have extensive lengths. This makes movie information search with sole aid of audio-visual data a daunting task. Therefore, movie search mostly relies on textual metadata associated with movies such as title, keywords, and plot.

In general, a movie includes certain textual information such as subtitles and annotations. In addition, movie scripts provide a wealth of information about the dialogue, characters, story, setting and action in movies. They are instrumental in bringing the filmmaker’s imagination to life. Most of the movie scripts conform to a set of semi-regular formatting conventions (www.screenwriting.info). These standard terms and conventions can be reaped in order to extract useful elements from scripts.

Scene boundary detection is a fundamental research area

which allows navigating movies with semantic ways. A movie consists of a thousand of shots and hundreds of thousands of frames. A scene is a set of shots and a boundary of one of these shots becomes a scene boundary. A number of studies have attempted scene boundary detection that mainly analyzes visual features and group shots according to the visual shot similarity [5], [6]. Scene boundary detection has also been attempted with the aid of scripts and subtitles. The script-subtitles alignment mechanism has been used in [7–9] for the purpose. It may be mentioned in this context that a script, unlike its subtitle counterpart does not convey any information about the time period in the movie.

Crafted aspects of movie dialogue are useful as they are authored deliberately to convey feelings and perceptions of the characters portrayed [2], [3]. Additionally, the screenplay specifies the emotion of utterances with psychological state descriptors. The current article describes a method of detecting scene boundaries/transitions from movie script dialogues. For the purpose, we have prepared textual resource containing structured information from movie scripts. Scripts available online have been utilized for the purpose. For the experiments, each script has been considered a sequence of dialogue units only. Then an attempt has been made to group the dialogue units scene-wise, as they would have appeared in the movie.

Our method segments the sequential dialogue units based on their changing semantic content and affect. It fragments the sequence of dialogue units in such a fashion that consecutive groups/clusters of dialogue text are most ‘dissimilar’, semantically and affect-wise. A scriptwriter frames his scenes carefully to narrate his story as per his intentions. We have assumed that each scene he writes is a coherent entity. Since we have chosen to use dialogue only, our notion of a scene for this task has been that a ‘scene’ is a dialogue sequence which sticks together semantically and sentiment-wise and it is somewhat different in the same sense from its neighboring ‘scenes’. We have formulated this task as an optimization problem and adopted a Genetic Algorithm (GA) approach to this head. Genetic Algorithms (GAs) are a class of adaptive, randomized search and optimization techniques guided by principles of natural evolutions and genetics [10], [11]. This being fundamentally an optimized segmentation task, GA seemed to be most appropriate for our purpose.

The rest of the paper is structured as follows. Section II provides the rudiments of movie scripts and a detailed description of resource preparation. Section III describes in

Manuscript received December 15, 2012. Manuscript accepted for publication January 11, 2013.

Amitava Kundu and Sivaji Bandyopadhyay are with the Department of Computer Science & Engineering, Jadavpur University, Kolkata-700032, India (e-mail: amitava.jucse@gmail.com, sivaji_ju_cse@yahoo.com).

Dipankar Das is with the Department of Computer Science & Engineering, National Institute of Technology, Meghalaya, Laitumkhrach, Shillong-793003, Meghalaya, India (email: dipankar.dipnil2005@gmail.com).

detail how Genetic Algorithm has been incorporated in our experiments. Section IV provides the results and observations. Finally, we summarize and present the future directions in Section V.

II. RESOURCE PREPARATION

Most of the feature-length films are produced with the aid of scripts [12] or screenplays. The script presents a detailed vision of the story, settings, dialogues, situations and actions of a film. It gives the filmmakers, the actors and the crews the necessary assistance in bringing the intended situation to life. Thus it can be looked upon as a rich resource that provides vivid textual descriptions of semantic objects within a film.

The Internet Movie Script Database (IMSDb; www.imsdb.com) is a huge online repository of movie scripts. Hundreds of movie scripts, coming straight from the filmmakers, are available for download. In the present task, scripts were collected from the above mentioned database but the collected scripts were subjected to some necessary preprocessing where the scenes were separately identified and stored in a structured format. The following subsection provides a general overview of the elements in a film script followed by the details of preprocessing steps.

A. Highlights of a Movie Script

As mentioned above, a movie script delineates the story, character, setting, action and dialogue of a film [9], [12]. Additionally, the camera directions and shot boundaries are also included. The script may go through a number of revisions prior to production of the movie. Often, after all the shots have been edited the resulting movie may not conform to all the minute details of the original script.

In general, the actual content of a script follows a semi-regular format as shown in Figure 1, which shows a snippet of script from the movie Godfather-II. In a script, a scene starts with the slug line which establishes the physical context of the action that follows. It indicates whether a scene takes place inside or outside (INT or EXT), the name of the location (e.g. ‘THE CHURCH PLAZA’), and can potentially specify the time of day (e.g. DAY or NIGHT). Important people and key objects are usually highlighted by capitalizing their names.

The bulk of a script is comprised of dialogue description. Actions of characters are described along with the dialogues for each character. Dialogues have the character name (usually in all caps) followed by the dialogue text. An optional (V.O.) or (O.S.) may follow the character name indicating that the speaker should be off-screen (V.O. stands for a ‘Voice Over’). In most scripts, dialogue actions are clearly separated from action and camera directions by enclosing the non-dialogue descriptions in brackets. Usually scarce, parenthetical directives convey phenomena such as inflection or mood. An example of a dialogue unit follows:

GARDNER: *(a little nervous)* I've heard a lot about you, Mr. Corleone.

```
EXT. THE CHURCH PLAZA - NIGHT

The men continue on their night-walk, up to the plaza of the
church.

                STROLLO
                (Sicilian)
                ...who harbors the boy Vitone
                Andolini.

The figure of a single man on a mule passes them.

                MOSCA
                (Sicilian)
                Let no one give help to the boy
                Vito Andolini...

The man on the mule makes his way out of the village and
disappears into the distance.

We begin to hear, very quietly, the Waltz repeated once again.
```

Fig. 1. Portion of a screenplay of the film Godfather-II

```
CONTEXT: Disgusted, she turns around, and heads toward the master.
DEANNA: Don't follow me!

SCENE: EXT. TAHOE LAWN AND TABLES - MED. SHOT - DAY
CONTEXT: Rushing through the tables, waving an arm jangling with gold jewelry,
CONNIE: Mama...Mama! Here I am!
CONTEXT: She throws her arms around her Mother, who returns the affection somewhat.
MAMA: Constanzia. We expected you last week; we sent the car to pick you up at the airport.
CONNIE: I know, it was chaos; but anyway, here I am one week late. (lifting a shiny green bag)
MAMA: (not giving him a chance to greet her) Yes, thank you.
CONNIE: How are the kids?
MAMA: Well, thank you, they asked for you all week.
CONNIE: I got surprises for everybody!
MAMA: (glancing at the wrapping) Bought at the airport.
CONNIE: (gazing about) This is swell. Where's Michael? I've got things to get straight with him.
MAMA: You go see your children first, and then you wait to see your brother like everybody else.

SCENE: EXT. THE BOATHOUSE - DAY
```

Fig. 2. Snippets of a resource file showing a typical scene from Godfather-II screenplay

```
DEANNA: Don't follow me!

CONNIE: Mama...Mama! Here I am!
MAMA: Constanzia. We expected you last week; we sent the car to pick you up at the airport.
CONNIE: I know, it was chaos; but anyway, here I am one week late. (lifting a shiny green bag)
MAMA: (not giving him a chance to greet her) Yes, thank you.
CONNIE: How are the kids?
MAMA: Well, thank you, they asked for you all week.
CONNIE: I got surprises for everybody!
MAMA: (glancing at the wrapping) Bought at the airport.
CONNIE: (gazing about) This is swell. Where's Michael? I've got things to get straight with him.
MAMA: You go see your children first, and then you wait to see your brother like everybody else.

PENTANGELI: Hey, kid! You got any red wine?
```

Fig. 3. Snippets of a resource file showing the dialogues of the scene from Godfather-II shown in Fig. 2.

B. Preprocessing of scripts

As mentioned already, movie scripts are written in a semi-regular format. Variety of formats makes preprocessing a challenging task. To accomplish our experiments, we needed to prepare a structured resource from scripts so that we could evaluate the success of our scene boundary detection algorithm later on. Using knowledge of the usual anatomy of movie scripts outlined in the above subsection, the

information pertaining to different scenes has been identified. In each of the sets of resource files, one for each script, the scenes have been stored in a structured format, as indicated in Figure 2. A scene sets off with the scene heading and the following line describes either a dialogue or an action. Each dialogue turn starts with the name of the speaker in capitals followed by the text. An action starts off with the word “CONTEXT” and describes what is going on in the scene. Each scene is separated from the following and the preceding ones by blank lines. A significant amount of manual labor has been invested in making necessary refinements and corrections before storing the resource in the required structured format.

Additionally, another set of resource files where scene-wise dialogues are stored have been prepared. Here dialogues from different scenes are separated with blank lines. All context information and scene headings have been excluded. These files have been prepared so that a script can later on be considered as a sequence of dialogue units without any knowledge of scene boundary (see Figure 3). However, scenes without dialogue were also abundant. They have been thus ignored altogether.

III. SYSTEM FRAMEWORK

In the experimentation, each script has been considered merely as a sequence of dialogue units excluding the context information (and even parenthesized directives).

Our objective was to detect the scene transitions looking at a sequence of dialogue texts. A close manual examination revealed abrupt changes in overall topic and affect across dialogues from consecutive scenes in many cases. This observation motivated us to take an attempt for detecting scene transitions. We have incorporated a genetic algorithm approach for the purpose. Figure 4 and Figure 5 illustrate the scheme and convey the objective.

A. The Genetic Algorithm Approach

Genetic Algorithm (GA) is a search heuristic that mimics the process of natural evolution. GAs belong to the larger class of Evolutionary Algorithms (EA), which generate solutions to optimization problems using techniques inspired by natural evolution: crossover, mutation and selection.

In the present work, the challenge is to frame the task of detecting scene boundaries as an optimization problem. As mentioned above, the changes in topic or overall affect of dialogues have often been observed across consecutive scenes. Therefore, we have made a principal assumption that dialogues from consecutive scenes are “dissimilar” in some sense. The main idea is to segment the sequence of dialogue units into scenes in such a way that makes the consecutive dialogue groups as “dissimilar” as possible. In order to implement GA, we have utilized the Java Genetic Algorithms and Programming Package (JGAP version 3.6.2), an open-source Java based tool.

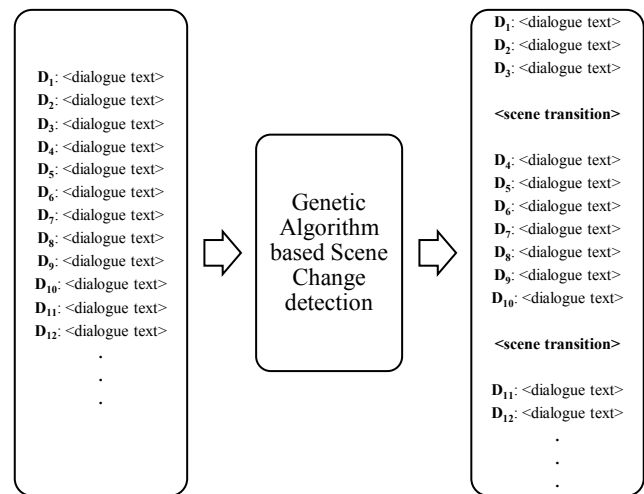


Fig. 4. The basic objective of our scheme. The input is a sequence of dialogue text as spoken in sequence by characters in film; D₁, D₂, D₃.... The system outputs a scene-wise grouping of dialogue units. A group can be seen as set of dialogue units belonging to a particular scene

1) Encoding.

As a genetic algorithm applies to encoded candidate solutions or chromosomes, before one can proceed, an encoding technique of chromosomes has to be chalked out. We have devised a binary encoding technique for the purpose where a chromosome is a string of ‘0’s and ‘1’s. As already mentioned, we have treated a script merely as a sequence of dialogue units. If we consider a script with n dialogue units, in our scheme, each chromosome is a binary string of length $n-1$. The encoding is such that a ‘1’ at m^{th} allele position of a chromosome indicates a scene transition after the m^{th} dialogue unit while a ‘0’ indicates scene continuation. Since the n^{th} dialogue unit would be the last one in the script, encoding a transition for it in the solution string would be unnecessary. Hence the length of such a string/chromosome is $n-1$. Figure 6 illustrates the scheme vividly.

2) Crossover, Mutation and Selection.

Since binary encoding has been adopted, usual genetic operators have been conveniently deployed. Single point crossover has been adopted. Mutation involves bit flipping where a ‘0’ is replaced by a ‘1’ and vice-versa. Standard Roulette Wheel Selection has been applied.

3) Population Size and Evolution.

We have incorporated the elitist model of GA wherein the fittest chromosome from each generation is preserved so as to ensure the best solution is not lost. Population size has been kept 50. We run the GA up to 100 generations.

4) Fitness Computation.

The fitness function is at the heart of our genetic algorithm framework. Dialogues across consecutive scenes have been assumed to be different or dissimilar.

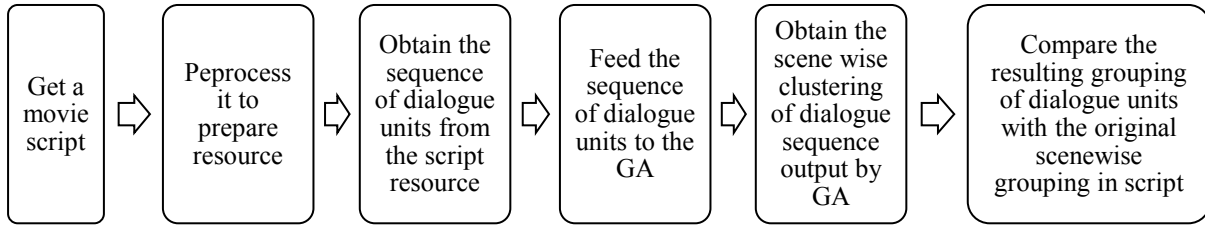


Fig. 5. Workflow of our system

In our scheme, those solutions that encode scene transitions resulting in greater dissimilarity across consecutive scenes have been considered as fitter chromosomes. Thus, given the binary encoding of each chromosome, the resulting scene-segmentation of dialogue units has been easily reconstructed. Thereafter, dissimilarity across consecutive groups of dialogue has been capitalized in computation of fitness. The SentiWordnet 3.0 [13] scores and Wordnet similarity of encoded scenes have also been accounted for.

(a) *SentiWordnet scores.*

For every script, the Stanford CoreNLP¹ (version 1.3.3) has been used to tokenize dialogue unit text, identify the named entities in turns and assign part-of-speech (POS) tags to tokens in a turn. Thereafter SentiWordnet 3.0 has been used to compute the total sentiment scores² of each dialogue unit. Sentiment scores of nouns excluding named entities, adjectives, adverbs and verbs only have been considered for this computation.

Consider the dialogue unit given below,

DEANNA: *Relax, Freddie honey. Come dance with me.*

The POS tagged version of the above with sentiment values indicated in parentheses is as follows,

*Relax*¹/**VB**(-0.06668) *,Freddie*/**NNP**honey²/**NN**(0.04166)
*,. Come*³/**VB**(0.00529) *dance*⁴/**NN**(0.0) *with*/**IN** *me*/**PRP** *,.*

Likewise, sum total sentiment score of all dialogues in every scene was computed, as encoded in any given chromosome. If a chromosome encoded scenes (groups of dialogue) S_1, S_2, \dots, S_n with total sentiment values P_1, P_2, \dots, P_n respectively, its fitness $F_{sentiment}$ is computed as follows,

$$F_{sentiment} = \sum_{i=2}^n (P_i - P_{i-1})^2$$

(b) *Wordnet distance of keywords.*

Each dialogue unit has been tokenized, named entities have been identified and POS tags have been assigned to tokens using Stanford CoreNLP.

1 WIDOW: (Sicilian) They've killed young Paolo! They've killed the boy Paolo! 2 I ... (scene change) 3 WIDOW: (Sicilian) Don Francesco. You murdered my husband, because he would 4 not bend. And his oldest son Paolo, because he swore revenge. But I have, is only 5 nine, and dumb-witted. He never speaks. 6 0 ... (scene continuation) 7 DON FRANCESCO: (Sicilian) I'm not afraid of his words. 8 0 ... (scene continuation) 9 WIDOW: (Sicilian) He is weak. 10 DON FRANCESCO: (Sicilian) He will grow strong. 11 WIDOW: (Sicilian) The child cannot harm you. 12 DON FRANCESCO: (Sicilian) He will be a man, and then he will come for revenge. 13 WIDOW: (Sicilian) I beg you, Don Francesco, spare my only son. He is all I 14 have. 15 WIDOW: (continuing) But I will kill you myself! (she lunges at the Mafia ch 16 0 ... (scene continuation) 17 MOSCA: (Sicilian) Our Friend promises misery to anyone who harbors the boy 18 Vito Andolini. 19 STROGO: (Sicilian, O.S.) Our Friend will be hard with any family who gives 20 misery to anyone who harbors the boy, Vito... 21 MOSCA: (Sicilian O.S.) ...misery to any family who harbors the boy, Vito... 22 FATHER: (Sicilian) Vito... We pray for you. 23 MOSCA: (Sicilian O.S.) ...Andolini... 24 STROGO: (Sicilian O.S.) Our Friend promises misery to any family...	1 WIDOW: (Sicilian) They've killed young Paolo! They've killed the boy Paolo! 2 I ... (scene change) 3 WIDOW: (Sicilian) Don Francesco. You murdered my husband, because he would 4 not bend. And his oldest son Paolo, because he swore revenge. But I have, is only 5 nine, and dumb-witted. He never speaks. 6 0 ... (scene continuation) 7 DON FRANCESCO: (Sicilian) I'm not afraid of his words. 8 0 ... (scene continuation) 9 WIDOW: (Sicilian) He is weak. 10 DON FRANCESCO: (Sicilian) He will grow strong. 11 WIDOW: (Sicilian) The child cannot harm you. 12 DON FRANCESCO: (Sicilian) He will be a man, and then he will come for revenge. 13 WIDOW: (Sicilian) I beg you, Don Francesco, spare my only son. He is all I have. 14 In the name of the Holy Spirit. I swear he will never be a danger to you. 15 0 ... (scene continuation) 16 WIDOW: (continuing) But I will kill you myself! (she lunges at the Mafia chieftain) 17 Escape, go! 18 1 ... (scene change) 19 MOSCA: (Sicilian) Our Friend promises misery to anyone who harbors the boy Vito Andolini. (he turns and shouts in the other direction) Our Friend promises misery to anyone who harbors the boy Vito Andolini.
Encoding of the above snippet of dialogues (the blank line indicates a scene transition):	10000000111100...

Fig. 6. Snippet of dialogue sequence from Godfather-II and its encoding.

Additionally, each of the dialogue units has been annotated with the CoreNLP parse annotator so as to obtain the Stanford semantic dependency relations [14] of each turn. Object relations have also been identified (if present) in each turn namely *doobj*, *iobj* and *pobj*. Thereafter, the dependent components in those relations have been collected. A dependent in such an object relation has been considered as a potential keyword or topic word of the piece of dialogue. However, a dependent is ignored if it appeared in a list of stop words. An example with the keywords thus identified is given below (numbered ones are the identified words in question)

CONNIE: *I got surprises¹ for everybody²!*

Wordnet distance between keywords across consecutive dialogue groups have been considered for fitness computation. The `getDistance()` method of RiTaWordnet³ library has been deployed for computation of semantic distance between two words in Wordnet hierarchy.

Say the distance of two word is given by $d(w_1, w_2)$. It is the absolute value of the aforementioned distance metric between two words w_1 and w_2 in the Wordnet ontology.

¹ <http://nlp.stanford.edu/software/corenlp.shtml>

² Online resource: <http://sentiwordnet.isti.cnr.it/code/SWN3.java>

³ <http://rednoise.org/rita/wordnet/documentation/index.htm>

TABLE 1.
PERFORMANCE OF SCENE BOUNDARY DETECTION

Movie	#Dialogue Units in Script	Actual #scene transitions in script dialogue sequence	With F_{baseline}		With $F_{\text{sentiment}}$		With F_{wordnet}		With F_{combined}	
			P_{GA}	Transitions correctly detected	P_{GA}	Transitions correctly detected	P_{GA}	Transitions correctly detected	P_{GA}	Transitions correctly detected
<i>Godfather-II</i>	1190	195	31.2%	69.3%	51.3%	46.7%	57.6%	47.1%	55.6%	48.7%
<i>Casino</i>	2478	452	41.7%	64.4%	52.3%	49.1%	51.7%	42.7%	52.7%	49.6%
<i>Citizen Kane</i>	1014	88	37.8%	73.3%	52.1%	45.4%	58.8%	38.7%	55.4%	38.7%
Average			36.9%	69.0%	51.9%	47.1%	56.0%	42.8%	54.6%	45.7%

It gives the minimum distance between any two senses for the two words in the WordNet tree (result normalized to 0–1). If a chromosome encodes scenes (groups of dialogue) S_1, S_2, \dots, S_n , with sets of keywords K_1, K_2, \dots, K_n respectively, then the fitness F_{wordnet} is given by

$$F_{\text{wordnet}} = \sum_{i=2}^n \sum_{\substack{x \in K_i, \\ y \in K_{i-1}}} d(x, y).$$

(c) Combined Effect.

To examine the combined effect of the above two distances we have also experimented with the following fitness function which mixes the effect of both.

$$F_{\text{combined}} = F_{\text{sentiment}} F_{\text{wordnet}}.$$

B. Evaluation of Solution

Once the fittest chromosome across all the generations was obtained, the solution was obtained. We have used the Hamming distance for the evaluation purpose as our solutions have been encoded using binary encoding. Hamming distance of two binary strings is the number of bit positions that differ. Since we had prepared our resource files with the scene-boundary information, we were able to construct the actual solution as per our encoding scene. Then we compared the solution obtained using the GA with the former. If S_{actual} was the binary string representing the actual solution and S_{GA} the solution obtained using GA, we defined the performance metric P_{GA} as follows:

$$P_{GA} = \left(1 - \frac{\text{HammingDistance}(S_{\text{actual}}, S_{GA})}{\# \text{Dialogue units in the script} - 1} \right) \times 100\%$$

$$= \frac{\# \text{Bit positions that match between } S_{GA} \& S_{\text{actual}}}{\text{Size of encoding of segmented dialogue sequence}} \times 100\%.$$

The present task is not a supervised learning problem. So accurate spotting of scene transitions is understandably difficult. Thus, besides counting the number of correct transitions located, we have also tried to evaluate the performance differently. The above metric measures the similarity of the dialogue grouping output by the system with the original scene-wise grouping.

We present a simple baseline for comparison. We do not use identities of speakers (their names) of turns anywhere in the experiments. Instead, word overlap of scenes was used in devising a baseline. Say, S_i and S_{i-1} are adjacent dialogue scenes with the sets of words (stop words and named entities excluded) W_i and W_{i-1} . The word overlap can be given by the Jaccard index:

$$J(S_{i-1}, S_i) = \frac{|W_i \cap W_{i-1}|}{|W_i \cup W_{i-1}|}.$$

The closer $J(S_{i-1}, S_i)$ gets to 1 the greater the similarity of adjacent scenes S_i and S_{i-1} . Put in another way, the closer $J'(S_{i-1}, S_i) = 1 - J(S_{i-1}, S_i)$ gets to 1 greater the dissimilarity of the same scenes in terms of common words. Thus the baseline fitness was devised as follows:

$$F_{\text{baseline}} = \sum_{i=2}^n J'(S_{i-1}, S_i)$$

IV. RESULTS AND OBSERVATIONS

A number of scripts have been chosen and processed for resource preparation. Scripts of three movies, namely *Godfather-II*, *Casino* and *Citizen Kane* have been chosen for the purpose of experimentation. These scripts have been picked up due to the presence of abundant dialogue units (in excess of 1000). Table 1 sums up the performance of the scene transition detection algorithm. It reports the proposed performance metric P_{GA} as well as percentage of transitions correctly detected.

When the encoded version of a solution is considered, the proposed P_{GA} reflects the extent of similarity of the obtained solution with the actual one. In other words it reflects a score for the obtained solution.

The results report an average accuracy of 47.1% in correct detection of scene transitions with the proposed fitness metric $F_{\text{sentiment}}$. Use of Wordnet features in fitness computation improves the average value of P_{GA} but results in a fall in average accuracy to 42.8%. The combined sentiment and Wordnet features produce an average accuracy of 45.7%. It seems that SentiWordNet features have been more informative than WordNet based distance.

We have observed a lot of false detections of scene boundaries (detection of a transition where the scene should actually continue), resulting in over-fragmentation of dialogue sequence. The modest values for P_{GA} may be attributed to the same. Moreover, no information about the distribution of number of transitions in scripts was available. Thus, the GA used in our experiments initializes each chromosome randomly without a bias to a '0' (scene continuation) or '1' (scene transition). There has been no bias in mutating a '0' or '1' valued gene.

As far as the baseline is concerned, its average P_{GA} score of 36.9% is outperformed although it gives higher average scene transition detection accuracy (69.0%). We observed that in the baseline system the number of scenes output by the system was way higher than that in the original scripts. For example, in *Citizen Kane* as many as 731 scenes were output while the script describes 88 scenes of dialogue only. In other words, the solution chromosome for all the three movies contain mostly '1's, under the binary encoding scheme. However, P_{GA} dips as expected.

V. CONCLUSIONS AND FUTURE DIRECTIONS

In this article, we have reported a mechanism to fragment a sequence of movie script dialogue into scene-wise groups. Wordnet based semantic relatedness and SentiWordnet based affective features have been fed to a genetic algorithm framework to obtain a fragmented, scene-segmented dialogue sequence. The initial results have been encouraging. As a part of future work we are planning to apply our mechanism to easily available movie subtitles. The time information present in subtitles may be capitalized. It may be noted that subtitles hardly contain information about the setting in the movie and conveys mostly the dialogue. The proposed scheme coupled with the time information of subtitles seems to be appropriate for locating scene transitions. Besides, the same algorithm can be suitably modified for identification of emotional episodes in conversation from subtitles or scripts. We are planning to

incorporate more textual features (e.g. co-reference resolution) into the proposed framework.

REFERENCES

- [1] J. J. Jung, E. You, and S. B. Park, "Emotion-based character clustering for managing story-based contents: a cinemetric analysis," *Multimedia Tools and Applications*, pp. 1-17, 2012.
- [2] G. I. Lin, and M. A. Walker, "All the world's a stage: Learning character models from film," In *Proceedings of the Seventh AI and Interactive Digital Entertainment Conference, AIIDE*, vol. 11, 2011.
- [3] M. A. Walker, G. I. Lin, and J. E. Sawyer, "An Annotated Corpus of Film Dialogue for Learning and Characterizing Character Style," *Proceedings of LREC*, 2012.
- [4] R. E. Banchs, "Movie-DiC: a movie dialogue corpus for research and development," In *Proc. of the 50th Annual Meeting of the ACL*, 2012, 2012.
- [5] M. Cooper, and J. Foote, "Scene boundary detection via video self-similarity analysis," *International Conference on Image Processing*, vol. 3, pp. 378-381, IEEE, 2001.
- [6] J. Wang, and T. S. Chua, "A cinematic-based framework for scene boundary detection in video," *The Visual Computer*, vol. 19, no.5, 2003, pp-329-341.
- [7] T. Cour, C. Jordan, E. Miltsakaki, and B. Taskar, "Movie/script: Alignment and parsing of video and text transcription," *Proc. 10th European Conf. Computer Vision*, 2008, pp. 158-171.
- [8] S. B. Park, H. N. Kim, H. Kim, and G. S. Jo, "Exploiting Script-Subtitles Alignment to Scene Boundary Detection in Movie," In *Multimedia (ISM), 2010 IEEE International Symposium on*, pp. 49-56. IEEE (2010)
- [9] R. Turetsky, and N. Dimitrova, "Screenplay alignment for closed system speaker identification and analysis of feature films," *Proc. IEEE Int. Conf. Multimedia and Expo (ICME'04)*, 2004, pp. 1659-1662.
- [10] D. E. Goldberg, "Genetic Algorithms in Search, Optimization and Machine Learning," Addison-Wesley, New York, 1989.
- [11] L. (Ed.) Davis, "Handbook of Genetic Algorithms," Van Nostrand Reinhold, New York, 1991.
- [12] A. Jhala, "Exploiting Structure and Conventions of Movie Scripts for Information Retrieval and Text Mining," *Interactive Storytelling*, 2008, pp. 210-213.
- [13] B. Stefano, A. Esuli, and F. Sebastiani, "Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining," In *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10), Valletta, Malta, May. European Language Resources Association (ELRA)*, 2010.
- [14] M. C. De Marneffe, C. D. Manning, "Stanford typed dependencies manual," nlp.stanford.edu/software/dependencies_manual.pdf, 2008.