

스토리텔링 콘텐츠의 효과적인 관리를 위한 영화 스토리 발단부의 자동 경계 추출

박승보
인하대학교 교육대학원
(molaal@estab.inha.ac.kr)

유은순
단국대학교 미디어콘텐츠연구원
스토리텔링 연구센터
(eunsoony@hotmail.com)

정재은
영남대학교 컴퓨터공학과
(j2jung@gmail.com)

.....

영화는 스토리를 전달하는 대표적인 매체 중의 하나이며 영화 속 스토리는 인물, 사건, 배경의 유기적인 연결을 통해 진행된다. 소설과 같이 스토리를 지닌 다른 매체와 마찬가지로 영화 역시 서사 구조를 갖고 있다. 즉 등장인물 간의 갈등과 해결이 인과 관계에 따라 전개된다. 영화의 서사구조는 아리스토텔레스의 3막 구조를 가지며 6단계로 세분화 될 수 있다. 전통적인 3막 구조는 발단부, 중반부, 결말부로 구성된다. 발단부는 영화에 등장하는 인물이나 배경을 소개하고 사건의 실마리와 갈등을 암시하는 장치들을 제시한다. 중반부는 내외적인 요인들에 의해 사건이 발전하고 갈등과 긴장을 고조시키는 부분이다. 결말부에서는 사건이 해결되면서 스토리의 주제 및 작가의 메시지가 전달된다. 특히 발단부는 스토리 진행을 위해 등장인물의 성격을 설정하고 배경을 지시하는 등의 다양한 정보가 노출되는 부분이기 때문에 영화의 축약이나 등장인물들의 중요한 정보를 추출할 수 있는 중요한 부분이다. 따라서 영화의 스토리를 추출할 경우 서사구조에 따라 가중치를 다르게 부여할 필요가 있다. 본 논문에서는 등장인물 수를 누적한 그래프를 이용하여 발단부와 중반부의 경계를 추출하는 방법을 제시한다. 발단부에서는 주요 등장인물이 소개되고 이들 간의 갈등이 암시되거나 사건의 실마리가 제시된다. 따라서 주요 등장인물의등장이 마무리되는 장면에서 일정 장면이 진행된 이후의 장면이 발단부와 중반부의 경계가 된다. 주요 등장인물은 주인공과 적대자, 그리고 주인공과 적대자를 보조해 주는 각각의 보조자(조연)로 구성되기 때문에 등장인물 수 누적 그래프에서 주요 등장인물의등장이 마무리되는 장면을 찾기 위해서는 단역을 제거하여야 한다. 또한 주요 등장인물이 더 이상 등장하지 않는 변곡점을 찾는 방법이 필요하다. 이를 위해 본 논문에서는 단역이 제거된 등장인물 수 누적 그래프를 그린 후 변곡점을 찾는 방법을 제안한다. 실제 발단부와 중반부의 경계는 주요 등장인물의 출현이 마무리 된 후 갈등과 사건의 실마리가 암시되는 추가적인 장면이 진행된 후에 나타난다. 따라서 다수의 영화를 통해 추가적인 장면의 개수를 설정하는 실험을 진행할 것이다.

.....

논문접수일 : 2011년 11월 18일 게재확정일 : 2011년 12월 19일
투고유형 : 학술대회우수논문 교신저자 : 박승보

1. 서론

영화는 영상과 음성에 대한 신호체계에 이루어진 데이터의 흐름이지만 인간이 인지하는 영화는

신호뿐만 아니라 영화 속에서 인식되는 객체들에 의해 묘사되는 스토리이다. 따라서 영화 분석은 영상과 음성과 같은 신호에 기반을 둔 분석 이외에 인간이 영화를 인식하는 인지 수준의 스토리에 대

* 이 논문은 2011년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No.2011-0017156).

한 추가적인 분석이 필요하다(Park et al., 2011). 이를 위해 스토리 정보를 추출하거나 스토리에 기반하여 영화를 분석하려는 다양한 연구들이 시도되고 있다(Jung et al., 2004; Riedl and Young, 2006; Weng et al., 2009; Park et al., 2011). 영화 속 스토리 분석은 인간의 인지차원의 정보를 파악하는 것으로 매우 어려운 작업이다. 하지만 영화의 스토리를 파악한다면 영화 브라우징이나 검색과 같은 영화 분석에 차원 높은 정보와 서비스를 제공할 수 있고 다양한 응용 분야의 성능을 향상시킬 수 있다. 스토리를 파악하기 위한 기존의 연구들은 다음과 같다.

- 스토리 모델 구축(Jung et al., 2004; Riedl and Young, 2006): 영화의 스토리를 모델링 하는 연구로 장면간의 유사도를 계산하거나 영화를 축약하는 작업을 하기 위한 선행 작업으로 활용된다.
- 등장인물 간의 관계 그래프를 기반으로 중심인물을 파악(Weng et al., 2009) : 장면간 속에 동시에 나타나는 인물들을 소셜 네트워크로 구성하여 등장인물들 중에 주연과 조연을 구별하는 연구이다.
- 시퀀스 기반의 분석 방법(Park et al., 2011) : 대화를 기반으로 등장인물간의 소셜 네트워크를 구성하여 중심성 분석과 서브 네트워크 분석을 통해 배역의 중요도와 시퀀스를 추출하는 연구이다.

하지만 이러한 연구들은 영화에서 나타나는 장면이나 대화의 위치를 고려하지 않는 문제점이 있다. 즉 영화의 발단부에 나타나는 영화 요소와 결말에 존재하는 요소들이 스토리적 영향력이 동일하다고 가정하는 것은 문제가 있다. 예를 들어, 영화에서 발단부에서 주인공이 등장하여 5초 동안 우는 장면과 결말에서 5초 동안 등장하여 우는 장면이

존재할 경우 두 장면이 동일한 시간과 등장인물을 보유하였다 할지라도 스토리에 미치는 영향력은 다를 수 있다는 것이다. 따라서 영화의 스토리 분석을 위해 서사구조에 대한 분석이 선행 되어야 한다.

영화와 같이 스토리를 갖는 콘텐츠는 서사구조를 중반부의 스토리 진행 단계를 얼마나 세심하게 나누냐에 따라 3단계부터 6단계까지로 구분할 수 있다. 3단계는 영화를 ‘발단부-중반부-결말’로 구분하며, 4~6단계는 3단계의 세 부분을 좀 더 상세하게 구분하여 단계를 나누는 것이다(조은하, 2008). 본 논문에서는 영화의 서사구조를 3막 구조에 기초하여 발단부를 찾아 내는 것을 목표로 한다.

발단부는 스토리 전개가 시작되는 지점이므로 등장인물의 소개나 배경의 제시, 사건의 실마리, 갈등의 암시를 제시한다. 예를 들어 발단부에서 등장인물의 유형이 프로타고니스트(protagonist)인지, 트리타고니스트(tritagonist)인지, 안타고니스트(antagonist)인지를 설명 한다(웰스 루트, 1997; 조은하, 2008). 따라서 영화를 요약하거나 스토리 정보를 추출할 때 발단부로부터 중요한 정보를 추출할 수 있다. 발단부는 스토리에 대한 전개보다는 스토리 전개를 위한 설정 단계에 해당되기 때문에 중반부와 결말에 비해 스토리 전개의 속도가 다르게 된다. 경우에 따라 하나의 대사를 통해서 등장인물의 많은 부분들을 설명할 수도 있다. 또한 스토리의 중요한 요소인 등장인물들의 감정량의 분포량도 매우 다를 수 있다. 감정량은 스토리상의 위기나 절정을 포함하는 중반부에 많이 분포한다. 따라서 발단부에서 발생하는 스토리와 중반부나 결말의 스토리를 구분할 필요성이 있다.

일반적으로 발단부는 등장인물들 중에서 주·조연에 해당하는 등장인물들을 출현시키고 성격을 설명한다(조은하, 2008). 일부의 경우에 조연이 중반부부터 출현하는 경우도 있지만 대다수의 영

화에서 발단부에서 주·조연이 소개되고 그들의 성격이 설명된다. 역으로 주·조연이 소개되고 설명되는 부분이 발단부라는 것을 이용하여 스토리의 발단부를 추출하는 것도 가능할 것이다 (박승보 외, 2011). 하지만 발단부에서 등장인물들을 소개하는 것 뿐만 아니라 추가적으로 갈등과 사건의 실마리를 암시하는 부분이 존재하기 때문에 등장인물들의 소개가 종료된 장면에서 추가적인 장면이 진행된 곳이 발단부의 경계가 될 것이다. 따라서 본 논문은 영화 시나리오에서 등장인물의 출현과 추가되는 장면의 수를 파악하여 스토리의 발단부의 경계를 추출하는 방법을 제시하는 것을 목적으로 한다.

이를 위해 본 논문은 다음과 같은 구성을 갖는다. 2장에서는 영화의 서사구조와 등장인물의 관계를 이해하기 위한 배경 지식과 영화 시나리오의 작성 형식에 대한 내용을 설명하고 3장에서는 영화 서사구조에서 발단부를 추출하기 위해 등장인물 누적을 이용한 발단부 추출 방법을 기술한다. 4장에서는 실험과 평가를 통해 제안하는 방법이 효과적인지를 기술하고 마지막으로 5장에서 연구에 대한 결론과 향후 연구 과제에 대해 기술한다.

2. 서사구조와 등장인물의 관계

2.1 서사구조의 단계

영화나 소설과 같은 스토리를 갖는 콘텐츠는 표 1과 같이 ‘전반부-중반부-후반부’의 3단계설, ‘발단-발전-절정-해결’ 또는 ‘발단-분규-절정-해결’의 4단계설, ‘발단-제1 행동 단계-제2 행동 단계-발발-결말’의 5단계설, ‘의도-의도 장애-의도 반전-위기-위기 반전-대단원’의 6단계설 등으로 분류된다(조은하, 2008). 3단계는 아리스토텔레스에 의해 제시된 분류방법으로 일반적으로 3막 구조로

불려진다. 본 논문은 3단계설을 적용하여 서사구조를 파악 하였으나, 전반부의 명칭을 보편적으로 많이 사용 되는 발단부로 통일하여 사용하겠다.

발단부는 전체적인 스토리 전개를 위해 기본적인 정보를 제공하는 부분이다. 등장인물을 제시하고 인물이 살아온 환경이나 사건이 벌어질 배경을 설정한다. 환경이나 배경은 시·공간적 묘사를 통해 설정된다. 스토리에 등장하는 주요 인물들이 묘사되고 나면 스토리 전개를 위해 갈등 발생을 위한 암시가 나타난다(박지훈, 2008; 조은하, 2008). 갈등은 우연하거나 인물의 성격적 이유로 인한 사건에 개입되면서 나타날 수 있는데 이 부분이 발단과 중반부를 나누는 전환점이 된다. 갈등 요소가 제시되기 위해 등장인물들과 그들의 성격이 설명될 필요성이 있다.

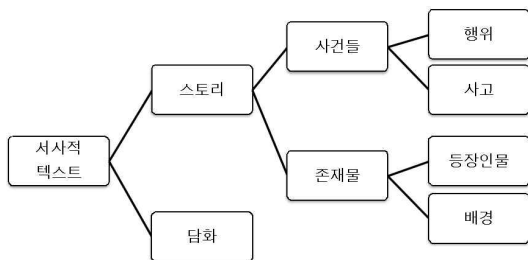
<표 1> 서사구조의 단계별 내용

3단	4단	5단	6단	내용
전반부	발단	발단	의도	등장인물의 소개 배경의 제시 사건의 실마리 제시 갈등의 암시
중반부	발전	전개	의도 장애	등장인물 성격의 변화·발전 다양한 복선과 암시
			의도 반전	사건의 본격적 전개 갈등과 분규의 제시
	절정	절정	위기	갈등과 분규의 상승 새로운 사건의 발생
위기 반전			갈등의 최고조 결말을 향한 분기점 해결의 실마리	
후반부	해결	결말	대단원	인물의 운명 결정 갈등과 분규의 해결

2.2 서사물의 구성요소

영화나 소설 등과 같은 서사적 텍스트는 구조주의적 관점에서 <그림 1>과 같이 스토리(histoire)와 담화(discours)의 두 부분으로 구분된다(김은

주, 정성환, 2007; 시모어 채트먼, 1990). 학자에 따라 서사, 스토리, 플롯, 서술행위(narration), 담화, 담론 등의 용어들이 혼용되어 사용되는 경우가 있다. 본 논문에서는 러시아의 형식주의에 따라 서사를 스토리와 담화로 구분한다(김은주, 정성환, 2007; 시모어 채트먼, 1990). 스토리는 텍스트가 말하고자 하는 내용을 의미하며 담화는 스토리를 표현하기 위한 형식으로 표현방식과 서술행위를 포함한다. 스토리는 행위와 사고에 의해 드러나는 사건들 및 영화에 등장하는 인물과 사건이 벌어지는 배경 등의 존재물로 구성된다.



<그림 1> 서사적 텍스트의 구성요소

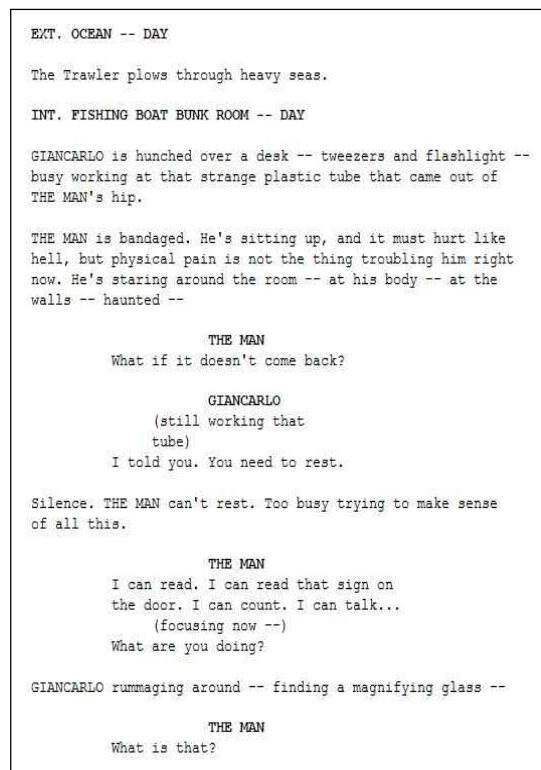
등장인물은 스토리 속 사건들을 일으키는 주체로써 스토리를 진행하는 중요한 요소가 된다. <그림 1>이 보여주고 있듯이 서사적 텍스트는 스토리 전개를 위해 등장인물을 필요로 한다.

2.3 Final Draft

시나리오는 스토리텔링을 위해 필요한 음성, 영상, 등장인물의 행동, 대사 등의 요소를 설명하는 문서이다.¹⁾ 영화 시나리오의 세계적인 표준 형식은 Final Draft 형식이다.²⁾ 이 형식은 미국의 할리

1) Screenwriting.info <http://www.screenwriting.info>.
 2) Final Draft, [http://en.wikipedia.org/wiki/Final_Draft_\(software\)](http://en.wikipedia.org/wiki/Final_Draft_(software)).

우드의 시나리오 작성의 표준으로 간주되고 있으며, 다른 나라에서도 이 형식을 준용하여 사용하고 있다. 시나리오는 일반적으로 'scene heading'(장면 머리글), 'action', 'character name'(등장인물명), 'dialogue'(대사), 'parenthetical', 'extension', 'transition', 'shot' 등으로 구성된다(Cour, 2008; Ronfard, 2003; Turetsky, 2004).



<그림 2> Final Draft 형태의 영화 시나리오의 예

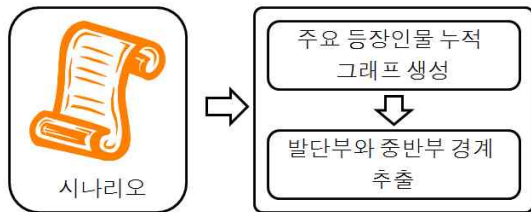
<그림 2>는 'scene heading', 'action', 'character name', 'dialogue', 'parenthetical' 등으로 구성된 영화 'The Bourne Identity' 시나리오의 일부분이다. 'scene heading'은 'INT' 또는 'EXT'로 시작한다. 이것은 장면의 시작을 표시하며, 촬영된 장소와 시간을 지시한다. 'action'은 장면이나 등장인물의 상

태를 설명한다. ‘character name’은 등장인물의 이름으로 ‘scene heading’ 내에 속하게 되며 이어서 등장인물이 말하는 대사인 ‘dialog’가 나타난다. ‘parenthetical’은 등장인물의 심리상태나 상황을 짧게 지시하는 부분이다.

영화의 시나리오에서 등장인물은 ‘character name’으로 장면 내에 명확하게 표현되며 대사를 하는 등장인물이 누구인 지를 명확하게 지시하게 된다. 본 논문은 등장인물의 누적을 위한 단위로 장면을 사용하였다. 장면은 ‘scene heading’을 사용한다. 다음 ‘scene heading’이 나타나기 전까지 존재하는 ‘character name’들을 한 장면에 등장하는 등장인물들로 사용하였다.

3. 등장인물 수를 이용한 발단부 경계 추출

본 논문에서 제안하는 시스템은 <그림 3>과 같이 이루어진다. 영화의 시나리오에서 발단부의 경계를 추출하는 방법은 크게 2단계로 구성된다. 2단계는 주요 등장인물 누적 그래프 생성 단계와 발단부와 중반부 경계 추출 단계로 구성된다. 첫 번째 단계인 주요 등장인물 누적 그래프 생성 단계에서는 시나리오에서 장면을 추출하고 신규 등장인물의 수를 장면의 진행에 따라 누적하여 표기하고 신규 등장인물들 중에서 단역을 제거한 주요

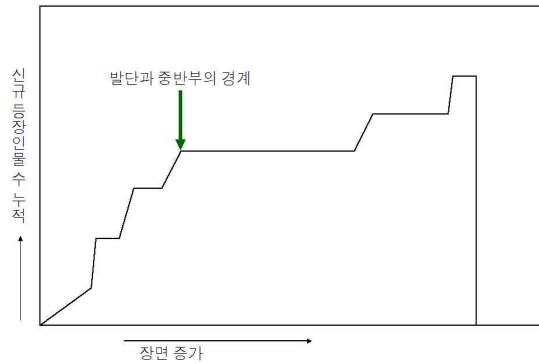


<그림 3> 발단과 중반부의 경계 추출

등장인물들만의 누적 그래프로 변환하는 작업들이 이루어진다. 두 번째 단계인 발단부와 중반부 경계 추출 단계에서는 주요 등장인물 누적 그래프에서 변곡점을 추출하고 일정량의 장면의 수를 합산하는 작업들이 이루어진다.

3.1 등장인물 수 누적 그래프

장면(scene)은 스토리를 구성하는 기본 요소로 본 논문에서 서사구조의 경계를 나누는 기본 단위로 사용되었다. 시나리오에서 장면을 추출하여 장면 내에서 출현하는 신규 등장인물들을 판단한다.



<그림 4> 등장인물 수 누적 그래프

발단에서 주요한 등장인물들이 소개되고 각각의 성격들이 설명되기 때문에 발단부의 후반부에서는 새로운 등장인물이 소개되는 장면은 감소될 것이다. 따라서 식 (1)에 의해 극의 진행에 따른 새로운 등장인물의 수를 누적하여 그래프로 작성하면 보편적으로 <그림 4>와 같은 그래프가 그려지게 된다.

$$Y(n) = \sum_{i=1}^n NC(i) \quad (1)$$

i : 장면 번호

$Y(n)$: n번째 장면에서의 등장인물 수 누적 값과 그래프

$INC(i)$: i번째 장면의 신규로 나타나는 등장인물 수

일반적으로 영화 초반에 새롭게 등장하는 인물 수는 급격이 증가하고 중반부터는 특별한 사건이 벌어지는 상황에서 국소적으로 증가하는 패턴을 나타낸다. <그림 4>에서 초반부의 증가가 멈추고 그래프가 평평해 지는 부분 이후에서 발단부와 중반부의 경계가 나타날 것이다. 하지만 중후반에 발생하는 국소적인 증가 구간으로 인해 오류가 발생할 수도 있다. 따라서 오류의 발생을 감소시키기 위해 추가적인 필터링이 필요하며 그래프의 기울기가 변하는 지점을 찾아내는 정확한 기준이 정해질 필요가 있다.

3.2 단역의 제거와 변곡점의 추출

영화에서 단역은 소수의 장면에 국한되어 등장하며 스토리 전개에 필요에 따라 중후반부에도 새로 추가되어 등장하는 경우가 빈번하다. 반면에 주연과 조연과 같이 스토리 전개를 담당하는 주요인물들은 갈등의 암시를 위해 영화의 발단부에서 대부분 소개 된다. 단역은 필요에 따라 극의 임의의 위치에서 등장할 수 있으므로 등장인물 수의 누적 그래프는 식 (2)와 같이 단역을 제거하여 주요 인물들만의 누적 그래프로 변경할 필요가 있다.

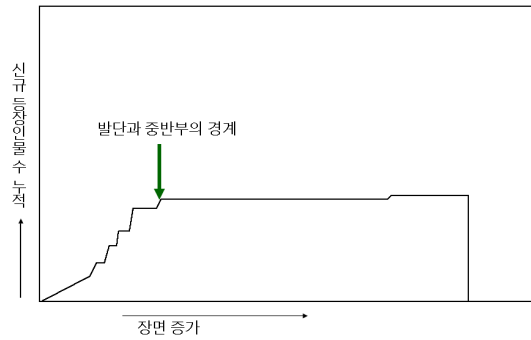
$$Y(n) = \sum_{i=1}^n INC(i) \quad (2)$$

i : 장면 번호

$Y(n)$: n번째 장면에서의 주요 등장인물 수 누적 값과 그래프

$INC(i)$: i번째 장면의 신규로 나타나는 주요 등장인물 수

등장인물 수 누적 그래프에서 단역에 대한 부분을 제거하면 <그림 5>와 같이 앞부분에서 주로 증가하면서 중후반에서는 수평이거나 증가량이 매우 미미한 정도의 그래프가 그려진다.



<그림 5> 단역이 제거된 등장인물 수 누적 그래프

신규 등장인물 수 누적 그래프를 그린 후 단역을 제거하기 위해 단역의 판단 기준을 정하였다. 등장인물들은 각자의 등장 장면수가 존재한다. 주·조연은 그 값이 높지만 단역은 매우 적은 값을 갖게 된다. 주·조연과 단역을 나누는 판단기준은 평균값으로 정하였다. 예를 들어 등장인물 A가 30장면, B가 20장면, C와 D가 각각 5장면일 경우 등장인물의 평균 장면 수는 $(30+20+5+5)/4 = 15$ 가 된다. 평균값이 15이며, 이 값보다 낮은 등장 장면수를 갖는 등장인물인 C와 D는 단역으로 분류 된다. 이렇게 각 영화마다 등장인물들의 평균 등장 장면수를 구하여 단역의 판단기준으로 사용하였다(Park, 2011).

발단부에서 주요한 등장인물의 소개가 끝나게 되므로 새로운 주·조연 급의 등장인물 출현은 종료 된다. 따라서 발단부를 찾기 위해 선행적으로 그래프의 기울기가 평평해지는 부분을 찾아야 한다. 이 부분 이후에 발단부와 중반부의 경계가 위치하게 될 것이다. 즉 <그림 5>에서 화살표로 표

시된 변곡점 이후에 발단부와 중반부의 경계 장면이 위치하게 된다.

이 지점을 시스템이 알아내기 위해서는 기울기가 급격하게 줄어 드는 지점을 알아 내도록 하면 된다. 기울기는 식 (3)과 같이 장면의 변화량에 대한 신규 등장인물 수 누적의 변화량이 된다.

$$IP = i, \text{ when } \frac{\Delta NC}{\Delta i} \leq \alpha \quad (3)$$

IP : 변곡점

ΔNC : 신규 주요 등장인물 수의 변화량

Δi : 장면의 변화량

기울기가 α 보다 작아지는 지점이 변곡점 IP 가 된다. 즉 식 (3)을 만족하는 i 값이 변곡점이 되는 것이다. 하지만 발단에서 새 등장인물은 장면의 진행에 따라 선형적으로 증가하는 것이 아니라 <그림 5>에서 보여지는 것처럼 계단식으로 증가한다. 따라서 장면 증가량은 1이 아닌 특정 범위를 선택하여야 한다. 대략 10장면 이상의 진행에 따른 주요 등장인물의 신규 출현을 고려하면 될 것으로 판단된다. 영화 시나리오의 총 장면수는 200개 이내에서 결정된다. 10장면은 영화 전체에 대해서는 5%의 진행을 의미하며, 발단부에 대해서는 단순히 계산하여도 16% 정도의 스토리가 진행되었다는 것을 담보하는 것이다. 10장면 이상으로 할 경우 일정량의 스토리가 진행되었다는 것을 담보할 수 있을 것이다.

3.3 발단부의 후반 장면

이후에 발단부가 끝나고 중반부가 시작되는 지점을 찾는 과정이 추가되어야 한다. 발단부는 주요 등장인물들을 소개한 후에 중반부와 결말에서 전개될 갈등과 사건 해결의 실마리를 제시하는 장면

들로 구성된다. 따라서 변곡점을 찾은 후에 일정량의 장면수를 더해줄 필요가 있다. 이를 위해 실험을 통해 평균적인 값을 찾아서 더해주는 작업을 진행하였다.

영화마다 전체 장면수가 다르기 때문에 비율적인 증가량을 계산하였다. 식 (4)와 같이 영화마다 전체 장면수에 대한 오차 장면 비율을 구하여 평균을 내어 변곡점의 장면번호에 더해주는 과정을 거쳤다.

$$SIP = IP + E \quad (4)$$

SIP : 발단부와 중반부의 경계 장면

E : 추가 장면수

4. 실험

본 논문에서 제안된 방법에 따라 발단의 경계를 추출하는 소프트웨어를 구현하였다. 소프트웨어는 마이크로 소프트 사의 윈도우 XP 운영체제 하에서 Embarcadero 사의 Delphi 7.0을 사용하여 개발하여 실험하였다.

실험과 평가를 위하여 <표 2>와 같이 영화 10편을 사용하였다. 실험은 영화의 시나리오들을 선택하여 시나리오 차원에서 발단 부분을 추출하여 시청자의 평가와 비교하였다. 영화의 시나리오는 The Internet Movie Script Database(IMSDB)³⁾에서 검색하여 사용하였다.

IMSDB의 시나리오들은 Final Draft의 형식을 따르고 있다. 시나리오에서 'scene heading'이 나타나면 새 장면이 시작된 것으로 판단하였으며 등장인물의 출현은 'character name'이 나타날 때로 하였다.

또한 변곡점을 찾기 위한 식 (1)의 α 값은 0.05로

3) <http://www.imsdb.com>.

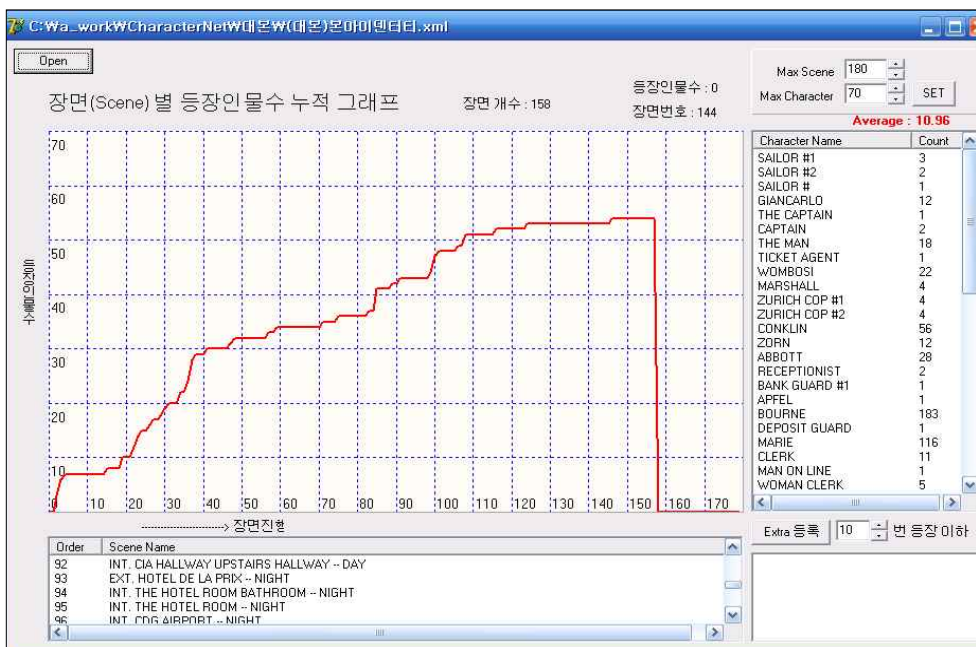
<표 2> 영화 데이터

ID	영화 제목	장르	전체 장면수
M1	The Bourne Identity(2002)	Action/ Crime/ Mystery	158
M2	New Moon(2009)	Adventure/Drama/Fantasy	121
M3	The Last Samurai(2003)	Action/Adventure/Drama	177
M4	The Back-up Plan(2010)	Comedy/Romance	133
M5	When In Rome(2010)	Comedy/Romance	79
M6	The Losers(2010)	Action/Crime/Drama	130
M7	Up(2009)	Animation/Adventure/Comedy	181
M8	Propose Day(2010)	Comedy/Romance	116
M9	The Departed(2006)	Crime/Drama/Thriller	207
M10	Notting Hill(1999)	Comedy/Romance	127

하였다. 이것은 20장면이 진행될 동안 1명의 신규 주요 등장인물이 발생하지 않는다면 발단에서 중반부로 전환되는 시점으로 판단한다는 것을 의미한다.

<그림 6>은 영화 M1의 등장인물 수 누적 그래

프를 개발한 프로그램으로 시나리오를 자동 분석하여 그려준 결과이다. 중앙에 그래프가 그려지며 우측에 등장인물 명들과 해당하는 등장 장면 수를 표시하였다. 이 영화의 경우, 중반부까지도 지속적으로 신규 등장인물이 나타나고 있어서 <그림 6>



<그림 6> M1의 등장인물 수 누적 그래프

의 그래프만으로 발단을 추출하기가 어렵다. 이의 개선을 위해 우측 하단에 단역을 제거하기 위해 단역의 명수를 지정하는 부분을 두었다.

단역의 판단 기준은 평균 장면수를 이용하였는데 <그림 6>의 등장인물 목록에서 계산된 평균 장면수를 이용하여 설정하였다. M1의 평균 장면수는 10.96이 나왔으므로 10이하의 등장 장면수를 갖는 등장인물은 단역으로 제거하였다. <그림 7>은 단역이 제거된 등장인물 누적 그래프 결과를 나타낸 것이다.

<그림 7>을 보면 단역이 제거되었기 때문에 영화의 초반부까지만 주·조연의 등장인물이 나타나는 것을 볼 수가 있으며 34장면 부근부터 신규로 주·조연이 등장하지 않아 그래프가 수평으로 유지되는 것을 알 수 있다. 따라서 본 논문은 수평으로 그래프가 유지되는 34장면에서 발단에서 중반부로 전환되는 것으로 판단할 수 있다.

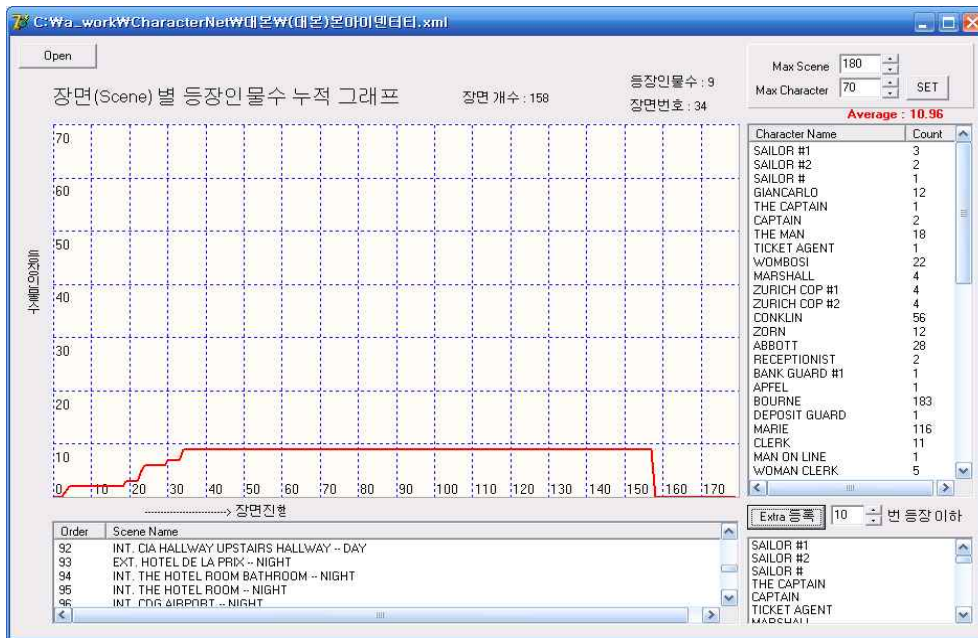
<표 3> 실험 결과의 평가

ID	자동 추출 결과	사용자 평가	오차(E)	오차율(E%)
M1	장면 34	장면 57	-23	14.56%
M2	장면 6	장면 25	-19	15.70%
M3	장면 31	장면 31	0	0%
M4	장면 20	장면 43	-23	17.29%
M5	장면 36	장면 18	+18	22.78%
M6	장면 33	장면 33	0	0%
M7	장면 49	장면 49	0	0%
M8	장면 26	장면 26	0	0%
M9	장면 49	장면 54	-5	2.42%
M10	장면 43	장면 38	+5	3.94%
평균(E /N)			9.3	7.67%

주) * N : 영화의 개수(여기서는 10개).

$$** E\% = \frac{\sum_{i=1}^{10} E}{\text{전체장면수}} \times 100$$

이러한 방식으로 주어진 10개의 영화를 분석하면 <표 3>과 같은 결과를 얻게 된다. <표 3>의 자



<그림 7> 단역이 제거된 M1의 등장인물 수 누적 그래프

동 추출결과와 사용자 평가의 값은 발단과 중반부의 경계가 되는 장면의 번호를 의미한다. 사용자 평가는 시청자에 의해 평가된 발단과 중반부의 경계장면이다.

각 영화들의 발단과 중반부의 경계 추출 성능의 오차는 <표 3>과 같이 나타나고 그 평균은 약 9.3으로 나타났다. 9.3의 오차는 자동 추출된 경계 부분이 평균적으로 9.3장면 벗어났다는 것을 의미한다.

8개의 영화에서 자동 추출결과가 사용자평가에 비해 적거나 같게 검출되었으나, M5와 M10 영화에서는 더 늦은 장면이 경계로 자동 검출되었다. 일반적으로 발단에서 등장인물의 소개가 이루어진 후에 갈등의 암시를 위한 추가적인 장면들이 배치되었다. 따라서 주조연의등장이 완료된 시점 후에 적당한 값이 추가되어 발단의 경계로 검출할 필요성이 있다. <표 3>의 오차율인 총 장면에 대한 7.67%의 장면수를 구하여 자동 추출된 결과에 더한 지점이 평균적으로 보다 더 정확한 발단부와 중반부의 경계로 추출될 수 있을 것이다.

5. 결론

본 논문은 영화의 서사구조에서 발단 부분을 자동으로 추출하는 방법을 제안하고 실험을 통하여 효과를 제시하였다. 10개의 영화에 대해 등장인물의 누적 그래프를 통한 경계추출 방법에 의한 결과는 평균 9.3정도의 오차를 나타냈다. 오차의 원인은 등장인물의 소개가 이루어진 후 갈등의 암시를 위한 추가적인 장면들이 존재하기 때문으로 판단되어 일정량의 오차 보정 값을 실험을 통해 찾아냈다. 영화에 존재하는 총 장면수의 약 7.67%를 더해줄 경우 보다 나은 경계추출 성능을 얻게 된다.

콘텐츠의 서사구조 파악은 영화를 축약하거나 영화로부터 정보를 검색할 때 고려해야 할 매우

사항이다. 또한 산업적으로 중요한 간접광고나 중 간광고의 효과적인 삽입을 할 때 활용될 수 있는 연구로 판단된다. TV 시청환경이 스마트 TV 환경으로 변화하는 현재의 시장상황에서 콘텐츠의 서사구조 파악을 통한 효율적인 광고시점 파악은 매우 적절한 연구가 될 것으로 기대한다.

본 논문은 영화의 발단 부분만을 추출하는 것을 목표로 하였으나 주요 등장인물의 등장 밀도나 주요 등장인물의 등장 패턴의 변화를 추적하면 중반부와 결말의 경계 또한 파악 할 수 있을 것으로 기대된다. 따라서 향후 영화의 서사구조 파악을 위한 등장인물들의 출현 패턴에 대한 추가적인 연구가 필요할 것으로 판단된다.

참고문헌

- 김은주, 정성환, “영상서사구조와 시간조작(Time Fabrication)에 대한 연구”, *디자인학연구*, 20권 5호(2007), 67~78.
- 박승보, 이예지, 유은순, 정재은, “등장인물 누적에 따른 스토리 발단 부분의 경계 판단”, 한국지능정보시스템학회 추계학술대회, 한국지능정보시스템학회, (2011), 135~140.
- 박지훈, *영화제작 매스터북*, 도서출판 책과길, (2008), 127~135.
- 시모어 채트먼, *영화와 소설의 서사구조*, 민음사, (1990), 17~28.
- 웰스 루트, *시나리오의 구성과 기법*, 현대미학사, 1997.
- 조은하, *콘셉트에서 스크립트까지 애니메이션 시나리오 쓰기*, 랜덤하우스코리아, 2008.
- Cour, T., C. Jordan, E. Miltsakaki, and B. Taskar, “Movie/script : Alignment and parsing of video and text transcription”, *Proceed-*

- ings of European Conference on Computer Vision, Part IV, (2008), 158~171.*
- Jung, B., T. Kwak, J. Song, and Y. Lee, "Narrative abstraction model for story-oriented video", *Proceedings of the 12th annual ACM international conference on Multimedia*, (2004), 10~16.
- Park, S.-B., K.-J. Oh, and G.-S. Jo, "Social Network Analysis in a Movie using Character-net", *Multimedia Tools and Application*, Online First, 2011.
- Riedl, M. O. and R. M. Young, "From Linear Story Generation to Branching Story Graphs", *IEEE Computer Graphics and Applications*, Vol.26, No.3(2006), 23~31.
- Ronfard, R. and T. T. Thuong, "A Framework for Aligning and Indexing Movies with their Script", *Proceedings of IEEE International Conference on Multimedia and Expo*, (2003), 1~21.
- Turetsky, R. and N. Dimitrova, "Screenplay Alignment for Closed-System Speaker Identification and Analysis of Feature Films", *Proceedings of IEEE International Conference on Multimedia and Expo*, (2004), 1659~1662.
- Weng, C. Y., W. T. Chu, and J. L. Wu, "Role-Net : movie analysis from the perspective of social network", *IEEE Transaction on Multimedia*, Vol.11, No.2(2009), 256~271.

Abstract

Extracting Beginning Boundaries for Efficient Management of Movie Storytelling Contents

Seung-Bo Park* · Eunsoon You** · Jason J. Jung***

Movie is a representative media that can transmit stories to audiences. Basically, a story is described by characters in the movie. Different from other simple videos, movies deploy narrative structures for explaining various conflicts or collaborations between characters. These narrative structures consist of 3 main acts, which are beginning, middle, and ending. The beginning act includes 1) introduction to main characters and backgrounds, and 2) conflicts implication and clues for incidents. The middle act describes the events developed by both inside and outside factors and the story dramatic tension heighten. Finally, in the end act, the events are developed are resolved, and the topic of story and message of writer are transmitted. When story information is extracted from movie, it is needed to consider that it has different weights by narrative structure. Namely, when some information is extracted, it has a different influence to story deployment depending on where it locates at the beginning, middle and end acts. The beginning act is the part that exposes to audiences for story set-up various information such as setting of characters and depiction of backgrounds. And thus, it is necessary to extract much kind information from the beginning act in order to abstract a movie or retrieve character information. Thereby, this paper proposes a novel method for extracting the beginning boundaries. It is the method that detects a boundary scene between the beginning act and middle using the accumulation graph of characters. The beginning act consists of the scenes that introduce important characters, imply the conflict relationship between them, and suggest clues to resolve troubles.

First, a scene that the new important characters don't appear any more should be detected in order to extract a scene completed the introduction of them. The important characters mean the major and minor characters, which can be dealt as important characters since they lead story progression. Extra should be excluded in order to extract a scene completed the introduction of important characters in the accumulation graph of characters. Extra means the characters that appear only several scenes. Second, the inflection point is detected in the accumulation graph of characters. It is the point

* Graduate School of Education, Inha University

** Storytelling Research Center, Institute of Media Content, Dankook University

*** Department of Computer Engineering, Yeungnam University

that the increasing line changes to horizontal line. Namely, when the slope of line keeps zero during long scenes, starting point of this line with zero slope becomes the inflection point. Inflection point will be detected in the accumulation graph of characters without extra. Third, several scenes are considered as additional story progression such as conflicts implication and clues suggestion. Actually, movie story can arrive at a scene located between beginning act and middle when additional several scenes are elapsed after the introduction of important characters. We will decide the ratio of additional scenes for total scenes by experiment in order to detect this scene. The ratio of additional scenes is gained as 7.67% by experiment. It is the story inflection point to change from beginning to middle act when this ratio is added to the inflection point of graph. Our proposed method consists of these three steps. We selected 10 movies for experiment and evaluation. These movies consisted of various genres. By measuring the accuracy of boundary detection experiment, we have shown that the proposed method is more efficient.

Key Words : Narrative Structure, Storytelling, Beginning, Character, Story Inflection Point

저자 소개



박승보

인하대학교 전기공학과를 1995년에 졸업하였고, 1997년에 인하대학교 전기공학과에서 석사학위를, 2011년에 인하대학교 정보공학과에서 박사 학위를 각각 취득하였다. 1996년부터 2002년까지 대우전자에서 주임연구원으로 재직하였으며, 2011년부터 현재까지 인하대학교 교육대학원에서 강사로 재직 중이다. 연구 관심분야는 비디오 스토리 분석, 스토리텔링 분석, 시맨틱 콘텐츠, 비디오 지식표현, 소셜네트워크 분석, 인공지능 등이다.



유은순

인하대학교 불어불문학과를 졸업하였고, 2001년과 2007년에 프랑스 Besancon University에서 언어학(세부전공 : 자연언어처리)석사 학위와 박사 학위를 각각 취득하였다. 2011년부터 현재까지 단국대학교 미디어콘텐츠연구원 스토리텔링 연구센터에서 전임연구원으로 재직 중이다. 연구 관심분야는 시맨틱 웹, 기계 번역, 온톨로지, 디지털 스토리텔링, 소셜네트워크 등이다.



정재은

현재 영남대학교 컴퓨터공학과에 조교수로 재직 중이다. 연구분야는 Description Logic과 Knowledge Engineering이다.