

등장인물 기반의 영화의 스토리 비교 방법론 연구

Research of Methodology to Compare Movie Stories

박 승 보, 김 현 식*, 유 은 순**
 인하대학교*, 단국대학교**

Park Seung-Bo, Kim Hyun Sik*, You Eun-Soon**
 Inha Univ. *, Dankook Univ. **

요약

사람이 영화를 이해하는 주된 내용은 스토리이다. 따라서 영화를 검색하거나 추천하기 위해서는 스토리 차원의 영화 분석이 선행되어야 한다. 더욱이 영화 추천이나 검색을 위해서는 영화간의 스토리차원의 비교를 수행할 수 있는 방법론에 대한 연구가 필요하다. 이를 위해 본 논문에서는 등장인물 기반으로 하는 영화 정규화 방법론을 소개하고 군집화를 통해 그 의미를 고찰한다.

1. 서론

현재의 내용기반의 영화 분석은 시청각적 데이터를 활용하는 연구가 주류를 이루고 있다. 시청각적 데이터를 활용할 경우 비교적 손쉽게 원하는 정보를 검색하는 것이 가능하다. 하지만 인간이 인식하는 영화의 핵심요소는 스토리이기 때문에 시청각적 데이터만으로는 검색에 한계를 갖는다. 시청각적 데이터만으로는 개체간의 관계나 등장인물의 중요도 등에 대한 고려를 하기가 어렵다. 이의 개선을 위해서는 정보 검색 시에 스토리에 대한 고려가 필요하다.

이를 위해 본 논문에서는 등장인물 기반의 스토리 모델링 기법인 Character-net 기반의 영화 정규화와 비교 방법을 제안하고 실험을 통해 의미를 파악한다. Character-net은 등장인물을 노드로 하고 등장인물들 간의 대화의 누적을 간선과 가중치로 하는 소셜 네트워크 형태로 표현된다. 영화의 스토리는 등장인물들이 어우러져서 만들어내는 일련의 사건들의 연속이다[1]. 스토리는 등장인물들 간의 관계로 형성되는 소셜 네트워크 형태로 표현될 수 있다.

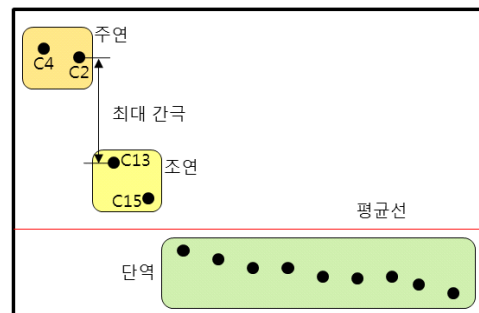
하지만 Character-net으로 형성된 영화에 대한 스토리 모델은 영화 간의 유사도 계산을 하려고 할 때 직접적으로 이용되기 어렵다. 그것은 영화마다 등장인물 수가 다르기 때문에 동일한 크기의 소셜 네트워크가 생성되지 않기 때문이다. 이를 위해 Character-net을 정규화 하는 방법이 필요하다. 본 논문에서는 Character-net을 주연과 조연과 단역으로 구성된 네트워크로 변환하여 정규화 하는 방법을 제안한다. 본 논문에서는 Character-net을 주연과 조연과 단역으로 구성된 네트워크로 변환하여 정규화 하는 방법을 제안한다. 본 논문에서는 Character-net을 주연과 조연과 단역으로 구성된 네트워크로 변환하여 정규화 하는 방법을 제안한다.

2. 정규화 된 Character-net의 구성

기존의 Character-net을 주연과 조연과 단역의 대표 값으로 이루어진 3 클래스 네트워크(3 class network)[2]은 스토리의 주요 특징들을 추출하여 대표 값으로 표현하는 정규화 과정이지만 영화간의 비교를 하기에는 특징 값들의 정규화가 추가로 필요하다. 3 클래스 네트워크를 중심으로 설명하면서 개선사항들을 추가하여 설명한다.

먼저 등장인물을 각 노드로 하는 Character-net을 주연

(Ma)과 조연(Mi)과 단역(Ex)의 3개의 클래스로 분류한다 [2]. 그림 1과 같이 전체 노드들의 연결정도 중심성에 대한 평균을 구한 후 그 이하는 단역으로 분류한다. 그리고 평균 이상의 노드들을 내림차순으로 정렬한 후 가장 큰 간극을 보이는 부분을 기준으로 큰 값의 노드들은 주연으로 작은 값의 노드들은 조연으로 분류한다[1,2].



▶▶ 그림 1. 등장인물의 분류

두 번째로 분류된 노드들의 연결정도 중심성(Degree Centrality, DC)[1]을 식 1을 이용하여 클래스 별로 합한다. 각각의 클래스는 연결정도 중심성의 합해진 값을 대표 값으로 하는 노드로 변환된다.

$$Value(Class_a) = \sum_{i=1}^n DC(N_i) \quad (1)$$

n : $Class_a$ 에 속한 노드들

N_i : $Class_a$ 에 속한 노드들 중 i 번째 노드

세 번째로 3개의 노드들이 연결될 수 있는 모든 간선을 생성하고 간선의 가중치를 식 2를 이용하여 지정한다. 가중치의 정규화를 위해 모든 가중치의 합으로 나누어 준다.

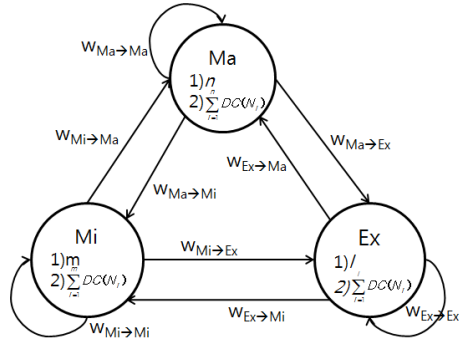
$$W(Class_a \rightarrow Class_b) = \frac{\sum_{i=1}^n \sum_{j=1}^m w_{ij}}{\sum_{k=1}^{\Omega} w_k} \quad (2)$$

n, m : $Class_a$ 와 $Class_b$ 에 속한 노드들

Ω : 모든 노드들

w_{ij} : $Class_a$ 노드에서 $Class_b$ 노드로 가는 간선의 가중치
 마지막으로 각 클래스에 속한 주연, 조연, 단역 노드들

의 개수 n, m, l 을 각각의 노드들의 대표 값으로 할당한다. 노드들의 개수는 정규화를 위해 전체 노드의 개수로 나누어준다. 결과적으로 생성된 정규화 된 Character-net은 그림 4와 같이 표현된다.



▶▶ 그림 2. 정규화된 3 클래스 네트워크

네 단계의 정규화 과정을 통해 Character-net으로부터 표 1과 같은 15개의 특징 값들이 추출되게 된다. 이 특징 값들을 활용하여 영화 간의 비교가 가능해 진다.

표 1. 정규화 된 3 클래스 네트워크의 특징값

순	특징값	내용
1	$Value(Ma)$	주연 노드들의 DC 값들의 합
2	$Value(Mi)$	조연 노드들의 DC 값들의 합
3	$Value(Ex)$	단역 노드들의 DC 값들의 합
4	$W(Ma \rightarrow Ma)$	주연과 주연 간의 가중치의 합
5	$W(Ma \rightarrow Mi)$	주연과 조연 간의 가중치의 합
6	$W(Ma \rightarrow Ex)$	주연과 단역 간의 가중치의 합
7	$W(Mi \rightarrow Ma)$	조연과 주연 간의 가중치의 합
8	$W(Mi \rightarrow Mi)$	조연과 조연 간의 가중치의 합
9	$W(Mi \rightarrow Ex)$	조연과 단역 간의 가중치의 합
10	$W(Ex \rightarrow Ma)$	단역과 주연 간의 가중치의 합
11	$W(Ex \rightarrow Mi)$	단역과 조연 간의 가중치의 합
12	$W(Ex \rightarrow Ex)$	단역과 단역 간의 가중치의 합
13	n	주연 노드의 개수
14	m	조연 노드의 개수
15	l	단역 노드의 개수

3. 실험 및 고찰

영화들 간의 비교를 위해 표 2와 같이 17편의 영화를 선정 하였다.

표 2. 실험 영화 데이터

ID (영화명)
M1 (Pretty Woman), M2 (Ugly truth-the), M3 (Spanglish), M4 (Propose-the), M5 (Lost in Translation), M6 (Bounty Hunter-the), M7 (Beginners), M8 (The Backup Plan), M9 (As good as it gets), M10 (500 DAYS OF SUMMER), M11 (Punch-Drunk Love), M12 (Ninotchka), M13 (Platinum blonde), M14 (Never been kissed), M15 (Larry Crowne), M16 (Its wonderful life), M17 (Crazy-Stupid-Love),

이들 영화에 대해 군집화를 통해 성능 평가를 진행하였다. 데이터마이닝 툴인 Weka를 활용하여 EM 군집화 방법을 적용하였다[3]. 군집은 표 3과 같이 2개를 지정하

여 실행하였다.

표 3. 실험 결과

군집	영화들
Cluster 0	M1, M2, M3, M4, M5, M6, M6, M7, M8, M9, M10
Cluster 1	M11, M12, M13, M14, M15, M16, M17

표 3에서처럼 2개의 군집으로 영화가 구분되었다. Cluster 0과 Cluster 1의 차이는 주연과 조연의 연결정도 중심성의 크기와, 주연과 조연들 간의 가중치 차이에 의해 구별되어졌다. Cluster 0은 주연들의 연결정도 중심성이 높으면서 주연들 간의 가중치가 높았다. 주연의 숫자도 상대적으로 높았다. 이에 비해 Cluster 1은 주연들의 연결정도 중심성이 낮고 조연들은 높았다. 주연들 간의 가중치는 Cluster 0에 비해 상대적으로 적게 나타났고 비교적 주연의 숫자보다 조연의 숫자가 많았다. 결과적으로 Cluster 0은 주연 중심의 스토리를 갖는 영화들의 군집이고, Cluster 1은 조연의 비중이 상대적으로 높은 군집이다.

4. 결론

본 논문에서 스토리 차원의 영화 비교를 위해 등장인물 기반의 스토리 모델인 Character-net을 이용한 영화 스토리 정규화 방법을 제안하고 영화들의 군집화를 통해 그 의미를 고찰하였다. 17편의 영화에 대한 군집화 실험 결과를 통해 주연 위주의 영화와 주조연 위주의 영화로 군집화 될 수 있다는 유의미한 결과를 얻었다.

하지만 군집화 시에 15개의 특징 값들 중에 주연과 조연의 값들이 주로 군집화에 영향을 미치는 것으로 나타났다. 향후 추가적인 연구를 통해 필요 없는 특징 값과 추가적인 특징 값들에 대한 고려가 필요할 것으로 판단 된다.

■ Acknowledge ■

이 논문은 2013년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임 (NRF-2013R1A1A2061737)

■ 참고 문헌 ■

- [1] Park, S.-B., Oh, K.-J. and Jo, G.-S., "Social Network Analysis in a Movie using Character-net," Multimedia Tools and Applications. Vol. 59, No. 2, pp.601-627, 2012, 7.
- [2] 박승보, 백영태, 유은순, "영화의 스토리 비교를 위한 Character-net 기반의 정규화 방법론," 제 25 회 영상처리 및 이해에 관한 워크샵(IPIU 2013), 한국방송공학회, pp.1-4, 2013.
- [3] Weka, <http://www.cs.waikato.ac.nz/ml/weka/>