Journal of The Korea Society of Computer and Information
Vol. 21 No. 5, pp. 19-29, May 2016

www.ksci.re.kr
http://dx.doi.org/10.9708/jksci.2016.21.5.019

Visualization of movie recommendation system using the sentimental vocabulary distribution map

Hyoji Ha*, Hyunwoo Han**, Seongmin Mun***, Sungyun Bae****, Jihye Lee*****, Kyungwon Lee******

Abstract

This paper suggests a method to refine a massive collective intelligence data, and visualize with multilevel sentiment network, in order to understand information in an intuitive and semantic way. For this study, we first calculated a frequency of sentiment words from each movie review. Second, we designed a Heatmap visualization to effectively discover the main emotions on each online movie review. Third, we formed a Sentiment-Movie Network combining the MDS Map and Social Network in order to fix the movie network topology, while creating a network graph to enable the clustering of similar nodes. Finally, we evaluated our progress to verify if it is actually helpful to improve user cognition for multilevel analysis experience compared to the existing network system, thus concluded that our method provides improved user experience in terms of cognition, being appropriate as an alternative method for semantic understanding.

▸ Keyword : Data Visualization, Semantic networks, Sentiment word analysis, Review data mining, Movie recommendation system, Sentiment movie network

Introduction

소셜 네트워크 분석(Social Network Analysis)은 네트워크가 가지는 고유의 구조 및 관계를 분석하여 사회적 기능 문제를 파악하고 해결하는 데 큰 역할을 한다. 따라서 데이터의 유사도를 기반으로 형성되는 네트워크 분석 및 사회 과학적 현상의 네트워크 분석, 그래프 이론, 추천시스템 등 광범위한 분야에서 활용되고 있다.

특히 네트워크 그래프를 그리는 대표적인 레이아웃 알고리즘인 Force-directed layout은 관련 있는 노드 간의 클러스터를 형성하게 함으로써 네트워크 분석을 위한 그래프를 그리는데 유용하다[1]. 그러나 Force-directed layout을 통해 그려지는 그래프는 노드 위치의 초깃값이 무작위로 설정되고, 노드 사이의 상대적인 관계에 따라 최종위치를 결정하기 때문에 데이터가 추가되거나 그래프를 새로 그릴 때마다 노드의 위치가 달라지는 문제점이 있다. 따라서 네트워크를 관찰하려는 사용자는 시스템에 대한 학습을 반복해야하는 불편함이 생긴다. (그림 1 참고) 이러한 문제점은 네트워크를 이루고 있는 데이터의 양이 많아질 때 네트워크 해석에 큰 장애요소가 될 수 있다. 또한, Force-directed layout을 그대로 적용하여 데이터를 시각화 한다면, 노드들의 위치가 달라진다는 문제로 인해서 집단 지성 정보의 의미 전달력을 상실할 가능성이 있다.

• First Author: Hyoji Ha, Corresponding Author: Kyungwon Lee
*Hyoji Ha(hjha0508@ajou.ac.kr), Life Media Interdisciplinary Program, Ajou University
**Hyunwoo Han(ainatsumi@ajou.ac.kr), Life Media Interdisciplinary Program, Ajou University
***Seongmin Mun(stat34@ajou.ac.kr), Life Media Interdisciplinary Program, Ajou University
****Sungyun Bae(roah@ajou.ac.kr), Life Media Interdisciplinary Program, Ajou University
*****Jihye Lee(alice0428@ajou.ac.kr), Life Media Interdisciplinary Program, Ajou University
******Kyungwon Lee (kwlee@ajou.ac.kr), Department of Digital Media, Ajou University
• Received: 2016. 04. 25, Revised: 2016. 04. 29, Accepted: 2016. 05. 12.
• This work was supported by the National Research Foundation of Korea Grant funded by the korean Government (NRF-2012S1A5A2A01020132).

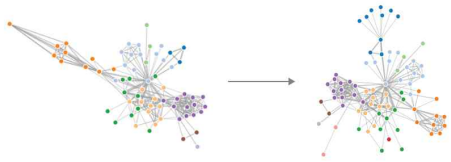

Fig. 1. Force-directed layout Network (Left: 40 nodes Right: 65 nodes): The location of nodes continues to change, whenever the data is added or modified

따라서 본 논문은 Force-directed layout이 가지는 문제점을 개선하며 네트워크를 의미적으로 해석할 수 있는 방법론을 제시하고자 하며, 방법론 적용의 모델이 되는 데이터를 집단 지성 데이터 중 하나인 '영화 리뷰 데이터'로 선정하였다. 정보시각화를 위해 다양한 분야의 데이터 중 영화 리뷰 데이터를 선정한 이유는, 네트워크를 의미적으로 해석하기 위해 다양한 의미 요소가 포함된 데이터가 필요했기 때문이다. 영화리뷰 데이터는 영화 리뷰어들의 평가 및 감정정보를 기반으로 다양한 해석이 가능하다는 장점이 있으며, 네트워크를 분석하는 사용자가 영화의 내용 및 분위기를 어느 정도 알고 있다면 네트워크의 성격을 쉽게 파악하고 공감을 할 수 있다는 장점이 있다. Force-directed 레이아웃이 가지는 문제점을 해결하기 위한 연구 방법을 서술하면 다음과 같다. 우선 리뷰 데이터에서 영화를 봤을 때 느낄 수 있는 대표적인 감정어휘 36개를 추출하였다. 추출된 감정어휘는 주성분 분석(Principal component analysis)을 이용하여 7가지 군집으로 분류하였다. 또한, 감정어휘간의 유사성 및 비유사성의 관계를 나타내기 위해 다차원 척도분석법(MDS: Multi-Dimensional Scaling)을 사용하여 감정어휘 간의 상관관계를 분석하였다. 그리고 상관관계 정보에 따라서 감정어휘 2차원 분포맵을 제작하였다. 그리고 감정어휘 2차원 분포맵을 기반으로 감정어휘 유사도 영화 네트워크를 구축하였는데, 네트워크 제작을 위해 두 가지 프로세스를 거쳤다. 우선, 네트워크를 이루는 노드(node) 하나는 하나의 영화정보를 포함하도록 만들고, 자신과 가장 유사한 감정을 가진 영화와 엣지(edge)를 구성하도록 하여 총 678개의 영화노드로 구성된 네트워크를 구축하였다. 또한, 영화 노드 하나가 가지고 있는 감정어휘 빈도 정보를 쉽게 파악하기 위해 히트맵 시각화(Heatmap visualization)를 적용하였다. 두 번째로, 영화 노드들이 감정어휘 2차원 분포맵 상의 의미적인 위치에 영향을 받도록 하여 노드의 절대적 위치가 노드가 가지는 감정어휘 정보를 반영하도록 하였다. 그 결과 각 영화를 나타내는 노드들은 감정어휘 빈도에 따라 2차원 분포맵 상에 있는 감정어휘의 공간적 위치에 이끌리도록 하는 네트워크 레이아웃이 만들어지게 되었고, 본 연구에서는 이를 '감정어휘 기반 영화 네트워크'라 명명하였다. 그리고 본 연구가 제안한 시각화 결과물이 영화를 추천과정에서 어떤 방식으로 사용되는지 보기 위한 샘플 시나리오를 제시하였다. 마지막으로 본 연구가 제안하는 시각화 방법론이 사용자들의 인지개선에 도움이 되는 지 검증하기 위한 실험을 실시하였다. 해당 실험에서는 히트맵 시각화의 적용 유무에 따라서 실험자가 네트워크를 탐색할 때 감정어휘 정보를 잘 이해할 수 있는 지를 비교하였다.

이상의 연구방법을 소개하기 위해, 본 논문은 '관련된 사전 연구 분석', '시각화에 사용될 데이터의 정제 과정', '두 개의 시각화 방법론 제안-히트맵 시각화, 감정어휘 기반 영화 네트워크', '시각화 검증 실험 및 통계 분석', '개발된 시각화 시스템을 이용한 영화추천 시나리오', '연구에 대한 결론 및 향후 연구 계획' 순으로 구성되었다.

II. Theory and previous research study

1. Related work

1.1 Sentiment words

감정언어에 대한 연구로 김명규[2]에서는 온라인상의 댓글에서 나타나는 감성단어 구축에 관한 시도가 이루어졌다. 이영희[3]의 연구에서는 사용자가 입력하는 단어에 따라 아바타가 반응하는 시스템을 위해 요구되는 감정 표현 어휘들을 수량화 이론 분석을 통해 분류, 분석이 이루어졌다. 감성어휘 공간을 나타내려는 연구로 성정연[4] 에서는 재질감을 표현할 수 있는 햅틱 형용사의 어휘를 도출하고 햅틱 형용사들의 관계를 다차원척도분석법으로 표현하였다.

1.2 Movie Recommendation

영화 추천 방법에 대한 연구는 크게 '정보 필터링 기술을 활용한 내용기반 추천 시스템'과 '협력적 추천 시스템'을 중심으로 이루어지고 있는데 그 중 내용기반 추천 시스템을 연구한 Oard[5]에 의하면 내용기반 추천시스템은 사용자의 개인정보를 기반으로 개인마다의 유형을 추출하여 이에 따른 선호도를 추정하는 것이 특징이었다. 협업 필터링을 통한 영화 추천 시스템 방식은 Sarwar[6], Adomavicius[7] 등에 의해 연구 되어 왔고, 사용자 정보와 유사한 정보를 가진 집단이 선택한 것들을 추천해 준다는 것이 특징이다. 두 연구 모두 사용자의 개인 정보를 이용하여 추천하는 방식을 선택하였는데, 본 연구에서 취급하고자 하는 데이터는 '영화를 보면서 느끼는 감정 리뷰 데이터'로써 사용자의 경험적 정보를 사용해 영화 추천 목적에 맞는 감정적 속성을 다양하게 반영할 수 있다는 것이 특징이다.

1.3 Network Visualization and Layouts

네트워크 시각화에 방법에 관한 연구는 다양하게 이루어져 왔는데, 그 중 Cody[8]의 연구에서는, 복잡한 형태의 네트워크 관계를 보다 쉽게 해석하기 위해 네트워크의 구조를 군집화하여 묶어서 간단히 나타내는 시각화 모델을 제시하였다. 그러나 이러한 방식은 사용자들이 단순화 된 네트워크 군집 안의 속성을 파악하지 못한다는 한계점이 존재한다.

네트워크 시각화에서 군집 관계파악의 모호함을 해소하고 가독성을 증가시키기 위한 연구로 Henry[9]의 연구가 있는데, 여기에서는 많은 양의 노드를 다루기 위해 연결 관계에 가장 중심이 되는 노드를 기반으로 타 노드를 군집화하여 제시한다. 그러나 노드의 복제를 통해 인위적으로 시각화를 왜곡시켜 사용자가 일정 크기 이상의 시각화를 분석할 때 복제된 노드를 구별하는데 혼선을 불러일으킬 수 있다는 점이 단점으로 작용한다.

III. Data processing and visualization

1. Data processing

1.1 Sentiment words collection

본 연구에서는 감정어휘의 분포맵을 제작하기 위해 한덕웅[10]의 연구를 참고하여, 834개의 정서용어 중에서 영화를 봤을 때 느낄 수 있는 감정어휘만을 분류하는 작업을 거치게 되었다. 이 작업을 위해 본교의 국어국문학과박사 전문가 1명과 본 연구를 진행하는 전문연구원 2명과 함께 서로 의견취합이 가능한 감정어휘만을 골라 최종 100개의 감정어휘를 선별하게 되었다. 다음으로 사용자들이 가장 많이 느끼는 감정어휘를 선별하기 위해 선정된 100개의 감정어휘를 토대로 설문조사를 실시하였다. 설문조사는 20대 대학생 30명을 대상으로 영화를 봤을 때 느낄 수 있는 감정에 대한 간단한 개념 설명을 거친 뒤에, 영화를 보는 상황일 때 해당 감정어휘를 느낄 수 있는 정도가 어떻게 되는 지를 조사하였다. 그 결과 평균이 상대적으로 낮은 감정어휘(4.00 '보통이다.'를 뜻하는 수치 이하) 32개를 추가적으로 제거하여 영화 추천에 적합한 68개의 감정어휘를 선정하게 되었다.

1.2 Sentiment Words Refinement

사용자 조사를 통하여 선정된 68개의 감정어휘 중에서 2차원 분포맵에 표현될 최종 감정어휘를 선별하기 위해 실제 영화리뷰에서 나타나는 감정어휘 데이터를 수집하여 비교하고, 리뷰에서 잘 나타나지 않는 감정어휘를 제거하는 작업을 시행하였다. 자세한 과정은 아래와 같이 3가지의 작업을 거쳤다.

1.2.1 Crawling

본 연구에서 사용한 영화리뷰 데이터는 한국에서 가장 많은 이용자를 보유한 포털사이트인 네이버의 영화정보 서비스[11]에서 수집하였다. 크롤링(Crawling)이란 웹페이지에서 데이터를 수집하는 작업과정으로, 본 연구에서는 영화리뷰의 감정어휘 수집을 자동화하기 위해 데이터를 수집할 수 있는 웹 크롤러를 제작하였다. 크롤러는 네이버 영화 홈페이지에서 특정 영화의 댓글과 리뷰들을 정제되지 않은 데이터 형태로 수집하는 단계와 수집된 데이터를 연구에서 사용 가능한 데이터로 가공하는 단계, 마지막으로 정제된 데이터를 분석하여 감정어휘를 추출해내는 단계로 설계되었다. 그 결과 2004년부터 2013년까지 한국에서 개봉된 2,289개 영화의 리뷰 4,107,605건이 수집되었다. 여기에서 2004년부터 2013년까지의 영화리뷰로 데이터의 범위를 한정지은 이유는, 네이버 영화 리뷰들의 댓글을 사전에 조사했을 때, 리뷰 댓글(감정어휘 형태소의 유무에 상관없이)이 1,000개 이상이 달린 영화들이 대부분 2004년~2013년 사이에 출시된 영화들이었기 때문이다. 2,289개의 표본을 기반으로, 감정어휘 빈도가 풍부한 개체만을 필터링하기 위해 감정어휘 형태소가 1,000개 이상이 있는 영화를 선별하였다. 그 결과 최종 678개의 영화 표본이 선정되어 네트워크 샘플 데이터로 활용되었다.

1.2.2 Establishing sentiment word dictionary

본 연구에서는 크롤링 작업을 통해 수집된 영화평들의 모든 텍스트 데이터들을 은전 한 닢[12]형태소 분석기를 사용하여 각각의 형태소들로 분리 하였다. 이 과정은 형태소 분석 작업 과정의 일환으로 수행되었으며, 형태소 분석이란 형태소를 비롯하여, 어근, 접두사/접미사, 품사(POS, part-of-speech) 등 다양한 언어적 속성의 구조를 파악하는 과정을 말한다.

형태소 분석 후 분리된 형태소들을 바탕으로 감정 형태소들을 추출하였으며, 선택한 감정 형태소들은 68개의 세부 감정어휘 카테고리에 각각 분류하여 감정어휘 별로 감정어 사전을 구축하였다. 감정 형태소들을 추출하고 카테고리들 안에 사전화하는 작업은 한국어 학자(한국어학 전공)의 자문을 받아서 진행하였다.

1.2.3 Applying TF-IDF(Term Frequency - Inverse Document Frequency)

본 연구에서는 실제 영화리뷰 데이터와 매칭과정을 통해 영향력이 미미한 감정어 집단을 제거하여 좀 더 정확한 결과를 얻고자 하였다. 이를 위해 우선 각 영화에서 각 감정어 집단(t)의 단어(w) 빈도수(tf: Term Frequency)를 구하였다.

$$tf(t,d) = \sum_{i=0}^{j} f(w_i, d) \qquad (1)$$

j= num of words in sentimental group t
The number of times that term t occurs in document d

그리고 역문서빈도(idf : Inverse Document Frequency)를 구하여서 보편적인 감정어 집단의 가중치가 낮아지도록 하였다. 각 영화에 대한 감정어 집단의 TF-IDF 스코어는 다음과 같이 구하였다.

$$idf(t,D) = \log\left(\frac{D}{d \in D : t \in d}\right) \qquad (2)$$

다음으로 감정어휘 개수를 줄이기 위해 각 감정어휘에서 나타날 수 있는 TF-IDF 스코어의 최대치를 구하였다.

$$TFIDF(t,d,D) = tf(t,d)*idf(t,D) \quad (3)$$

예를 들어 '경악하다'의 경우 모든 영화에서 TF-IDF 스코어의 비율이 0.8% 이하이다. 반면에 '달콤하다'의 경우는 적어도 한 개의 영화에서는 TF-IDF 스코어의 비율이 42%에 달하는 것을 뜻한다.

본 연구에서는 TF-IDF스코어의 비율이 10%미만인 감정어휘를 제거하고 최종적으로 36개의 감정어휘를 선택하였다. 선택된 36개의 감정어휘는 크게 Happy, Surprise, Boring, Sad, Anger, Disgust, Fear의 성격으로 나뉘게 되며, 감정어휘는 아래의 Table 1과 같다.

Table 1. 36 Sentiment Words.

Clustering Characteristics	Sentiment Words
Happy	Happy, Sweet, Funny, Excited, Pleasant, Fantastic, Gratified, Enjoyable, Energetic
Surprise	Surprised, Ecstatic, Awesome, Wonderful, Great, Touched, Impressed
Boring	Calm, Drowsy, Bored
Sad	Pitiful, Lonely, Mournful, Sad, Heartbroken, Unfortunate
Anger	Outraged, Furious
Disgust	Ominous, Cruel, Disgusted
Fear	Scared, Chilly, Horrified, Terrified, Creepy, Fearsome

2. Visualization

2.1 Heatmap visualization

히트맵 시각화는 각 타일이 색상 눈금을 가지고 그 색조로 값을 나타내는 시각화로써[12], Robert[13]의 연구처럼 데이터 행렬의 변칙 또는 패턴을 찾거나, Jeong[14]의 연구처럼 범위를 파악하는 연구에 사용되고 있다. 본 연구에서는 네트워크를 이루는 각 영화 노드들의 감정 분포도를 시각화하기 위해, 각 감정 어휘의 다차원 척도 분석 2차원 분포맵의 좌표 공간을 활용하여 감정 어휘의 단어빈도- 역문서 빈도의 크기를 히트맵 형태로 나타내었다. 우선, 데이터 정제과정에서 최종 선정된 36개의 감정어휘들 간의 거리를 측정하여 상관관계를 분석한 다음 다차원척도분석(MDS)을 실시하였다. 우선, 디지털미디어 및 영상을 전공하는 대학생 20명을 대상으로 36개의 감정어휘에 대해 의미상 거리 설문조사를 실시하였는데, 설문조사는 가로축 세로축 36개의 감정어휘를 배치한 설문지를 만들고 (36X36) 감정어휘간의 거리가 가장 가깝다고 느껴지면 3점, 가장 멀다고 느껴지면 -3점을 주는 방식의 리커트척도(Likert scale)를 이용하여 점수를 부여하는 형식으로 구성하였다. 20명이 기록한 데이터를 바탕으로 다양한 네트워크 분석기법이 활용 가능한 UCINET 프로그램을 사용하였고 이를 통해 영화 리뷰 감정어휘들 간의 의미상의 거리 기반으로 그림 2와 같은 Metric MDS를 형성할 수 있었다[15].

그 결과 X축의 양의 방향으로는 긍정적인 느낌을 가지는 "Happy", "Surprise" 와 관련된 군집 성격이 분포되었으며, X축의 음의 방향으로는 부정적인 느낌을 가지는 "Anger", "Disgust" 와 관련된 군집 성격이 분포되었다. 그리고 Y축의 양의 방향으로는 동적인(감정을 느낄 때 비교적 큰 제스처를 취할 수 있는)느낌을 가지는 "Fear", "Surprise"와 관련된 군집 성격이 분포되었으며, Y축의 음의 방향으로는 정적인(감정을 느낄 때 비교적 작은 제스처를 취할 수 있는)느낌을 가지는 "Sad", "Boring"과 관련된 군집 성격이 분포되었다. 그리고 2차원 분포맵에서 각각의 감정어휘들이 Happy, Surprise, Boring, Sad, Anger, Disgust, Fear 등의 성격에 따라 뚜렷하게 군집이 되는 것을 볼 수 있다. (그림 2 참고)

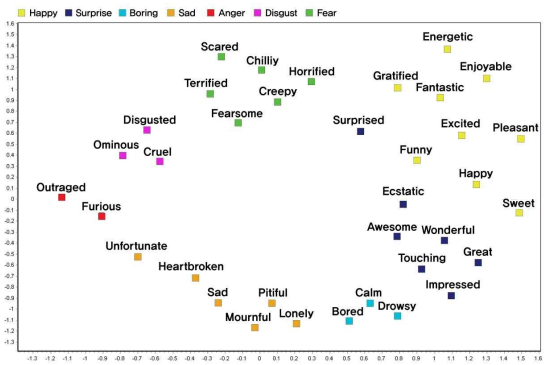

Fig. 2. 36 Sentiment words MDS Map

다음으로 2차원 감정어휘 분포맵을 기반으로 하여 히트맵 시각화를 제작하였는데, 히트맵을 생성하기 위해서는 임의의 영화 하나에 대해서 36개의 감정어휘(2차원 분포맵을 구성하고 있는 감정어휘들)에 대한 빈도수가 필요하다. 본 연구에서는 데이터 구축 과정을 통해 얻은 감정어휘 영화 리뷰데이터와 감정어휘 형태소 사전을 대조하여 각 영화에서의 감정어휘 빈도수를 측정하였다. 또한, 영화의 대표적인 성격과 관계없이 자주 등장하는 특정 감정어휘의 가중치를 낮추기 위해 단어빈도-역문서 빈도를 계산하여 수치를 조정하였다. 다시 말해 최종적으로 구해진 각 감정어휘의 단어빈도- 역문서 빈도가 해당 영화의 히트맵 시각화 그래프에 반영되는 실질적인 수치라고 할 수 있다. 최종 히트맵 그래프는 감정어휘의 2차원 분포맵을 배경으로 하고, 사각형의 작은 셀(cell)로 구성된다. 모든 셀은 0의 수치로 초기화되어 있으며, 해당 셀에 위치한 감정어휘 단어빈도- 역문서 빈도에 따라 초기화된 수치가 증가한다. 셀이 가지고 있는 수치가 높아질수록 다른 색으로 변함으로써 해당 감정어휘 단어빈도- 역문서 빈도의 높고 낮음을 확인 할 수 있다. 또한, 수치가 올라간 셀은 주위 셀의 영향을 미침으로써 히트맵

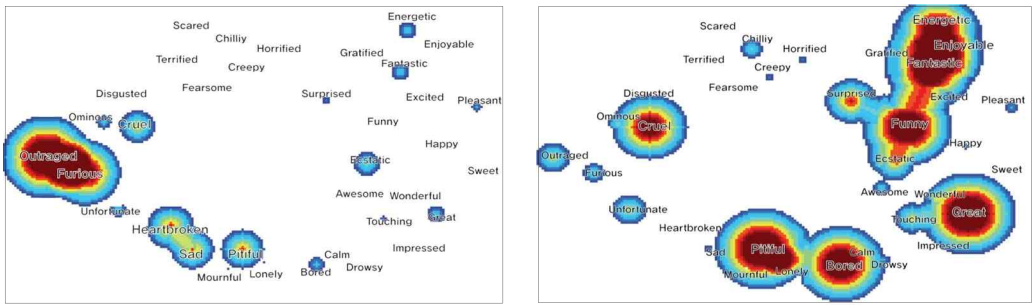

Fig. 3. (a) Heat Map of 'Don't Cry Mommy' which shows single emotion (Furious, Outraged) (b) Heat Map of 'Snowpiercer' which shows various emotions (Cruel, Pitiful, Lonely, Bored, Funny, Great and Energetic)

시각화의 모습은 지형도의 모습을 띠게 된다.

그림 3(b)는 영화 '설국열차(Snowpiercer)'에 대한 관람객들의 영화리뷰에 나타나는 감정어휘의 분포도를 히트맵 시각화로 나타낸 그래프이다. 그래프를 보면 관객들은 재미있고 대단하다(Funny and great)는 반응을 보이는 가운데 안타깝고 지루하다(Pitiful and Boring)는 감정 또한 높은 빈도를 보이고 있다. 실제로 영화 리뷰 중 하나의 샘플을 살펴보면 "중반까지 연출력이 돋보이는 작품이었다. 특히 횃불을 가지고 달려오는 장면은 봉준호 감독만의 느낌과 연출이 가장 돋보이는 장면이었다. 그러나 영화는 급격하게 지루해지고 영화를 지탱하던 긴장감마저 사라진다. 열차를 탐방하는 장면이 계속되고 허무한 결말로 끝이 난다."와 같이 영화에 대해 실망한 관객들이 있는 것을 볼 수 있어서 다양한 감정이 나타남을 알 수 있다.

실제로 영화 리뷰 중 하나의 샘플을 살펴보면 "중반까지 연출력이 돋보이는 작품이었다. 특히 횃불을 가지고 달려오는 장면은 봉준호 감독만의 느낌과 연출이 가장 돋보이는 장면이었다. 그러나 영화는 급격하게 지루해지고 영화를 지탱하던 긴장감마저 사라진다. 열차를 탐방하는 장면이 계속되고 허무한 결말로 끝이 난다."와 같이 영화에 대해 실망한 관객들이 있는 것을 볼 수 있어서 다양한 감정이 나타남을 알 수 있다.

영화의 감정어휘를 나타내는 히트맵 시각화는 크게 두 가지의 유형으로 나타난다는 것을 알 수 있었는데, 하나는 Happy, Surprise, Boring, Sad, Anger, Disgust, Fear등의 감정어휘 성격 중 하나의 성격에 대해서만 높은 빈도를 나타내는 경우가 있으며 (그림 3(a)) 다른 하나는 두 개 이상의 감정어휘 성격이 높은 빈도를 나타낸다는 것을 확인 할 수 있었다. (그림 3(b)) 두 개의 유형 중 두 번째 케이스를 더 많이 찾아볼 수 있었으며, 이를 통해 사람들이 영화를 볼 때 단일 감정이 아닌 두 개 이상의 복합된 감정을 느끼는 경우가 많다는 것을 히트맵 시각화를 통해 살펴볼 수 있었다. 또한, 서로 상반된 감정어휘들을 가지는 영화나 유사한 감정어휘들을 가지는 영화 노드들을 히트맵 시각화의 분포도 비교를 통해서 쉽게 이해할 수 있었다.

2.2 Sentiment-movie network

본 연구에서는 히트맵 시각화로써 영화 정보가 가지고 있는 감정어휘의 빈도를 나타내는 작업 이외에도, 감정어휘에 따른 영화 간의 유사도에 따른 네트워크를 제작하고자 한다. 이를 위해 영화리뷰의 감정어휘 분포맵 기준으로 영화 네트워크의 위상을 고정시킴으로써 노드의 수가 변경되어도 일정한 영역에 위치하게 되어 네트워크 구조를 쉽게 파악할 수 있도록 하였다. 본 연구에서는 이 시각화를 감정어휘 기반 영화 네트워크라 명명하였고, 그림 4는 감정어휘 기반 영화 네트워크의 기본 구조를 나타낸 것이다.

그림 4와 같이, 우리가 제안한 그래프는 두 개의 층으로 이루어져있다. 첫 번째 층은' 감정어휘 층'이라고 부르며, 36개의 감정어휘 2차원 분포맵으로 구성되어 있다. 감정어휘의 의미적인 위치는 초기 설정된 값에 위치하게 되며 고정된 상태를 유지한다.

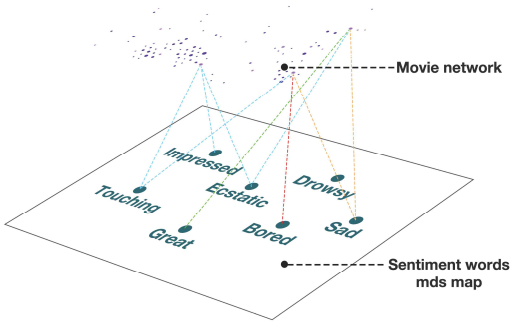

Fig. 4. Basic Structure of the Sentiment Movie Network

두 번째 층은 '네트워크 층'이라 부르며, 영화 네트워크를 구성할 노드들이 포함된다. 각각의 영화 노드는 유사도에 따라 다른 영화 노드들과 엣지를 형성하는 동시에, 해당 노드가 내포하고 있는 감정어휘에 따라서 감정어휘 2차원 분포맵의 감정어휘와 가상의 엣지를 형성한다. 그리고 Forced-directed algorithm에 따라, 엣지로 연결된 노드는 인력과 척력이 함께 작용한다. 반면에, 감정어휘의 의미적인 위치는 고정된 상태이므로 감정어휘로부터의 인력만 작용한다.

노드 간의 엣지 구성을 위해 36개 감정어휘의 단어빈도- 역문서 빈도를 기준으로 영화 간의 코사인유사도를 계산하였다.

또한, 노드와 감정어휘 사이를 연결해주기 위해서 감정어휘 빈도의 고정된 역치값(Threshold)을 설정하고 그 값을 초과하는 감정어휘 부분들에 대해서 노드가 힘을 받도록 하였다. 영화 A와 B의 유사도 SIM(A, B)를 계산하는 공식은 다음과 같다.

$$SIM(A,B) = \frac{\sum_{i=0}^{n} A_i * B_i}{\sqrt{\sum_{i=0}^{n} (A_i)^2} * \sqrt{\sum_{i=0}^{n} (B_i)^2}} \qquad (4)$$

이상의 공식에 따라서 나온 유사도를 바탕으로 본 연구에서는 감정어휘에 기반을 둔 네트워크를 형성할 수 있었고, 네트워크에서 노드가 위치하는 원리는 그림으로 설명하면 그림 5및 그림 6과 같다.

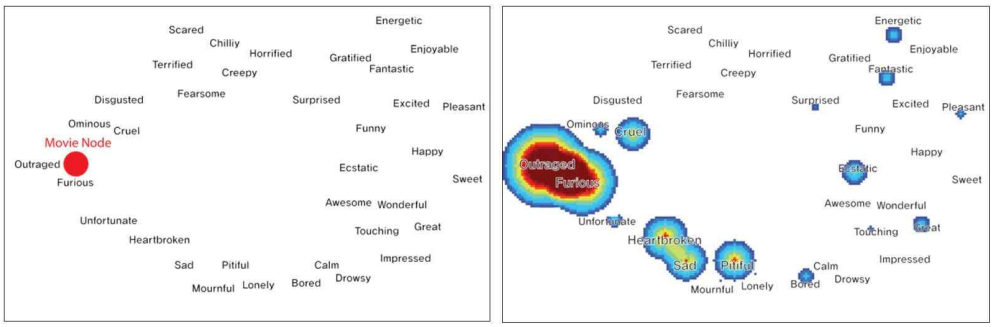

Fig. 5. Heatmap Visualization and positioning on the Sentiment-Movie Network (One point position) in case of "Don't Cry Mommy"

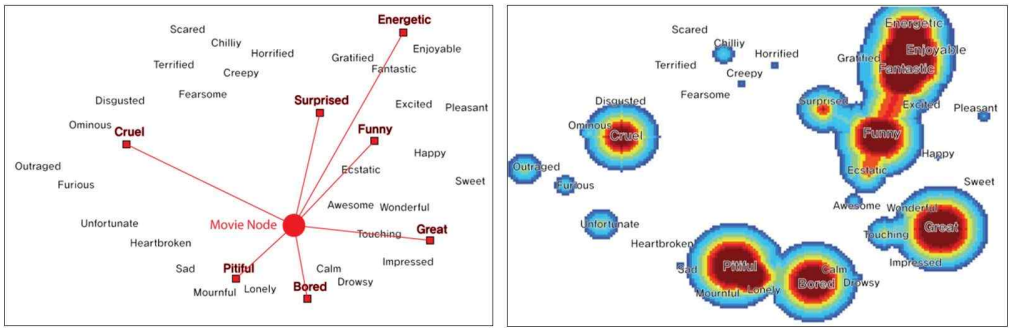

Fig. 6. Heatmap Visualization and positioning on the Sentiment-Movie Network (More than two point position) in case of "Snowpiercer"

그림 5과 그림 6은 히트맵 시각화를 통해 알 수 있는 감정어휘의 빈도에 따라 그래프 상에서 노드의 위치가 달라지는 예시를 보여준다. 그림 5의 경우 압도적으로 높은 빈도수를 가진 감정어휘 위치에 노드가 놓이는 것을 확인할 수 있다. 그림 6의 경우 높은 빈도수를 보이는 몇몇 감정어휘들에 의해 힘을 받기 때문에 결과적으로 2차원 분포맵 중간지점에 노드가 위치하는 것을 확인할 수 있다. 이와 같은 방법으로 네트워크로 연결된 노드들을 그래프 상에 위치시키면 영화 간의 연결성과 관련 감정어휘와의 연결성을 모두 고려하여 감정어휘의 빈도가 높은 공간에 유사한 영화끼리 군집을 형성한다.

마지막으로 각 노드군집들의 성격을 구분하기 위하여 코사인 유사도 값을 이용한 k-평균 알고리즘 군집화 작업을 하였다. k-평균 알고리즘 군집화를 위해, 각 노드들을 코사인유사도 기준으로 2차원 상의 분포시킨 뒤 임의로 k개의 군집으로 나눈다. 그 다음 각 군집의 무게 중심을 구하여 각각의 노드들을 각 군집의 무게중심 가운데 제일 가까운 것에 속하게 함으로써 새롭게 군집을 생성한다. 이 작업을 반복하다보면 노드들이 더 이상 소속된 군집을 바꾸지 않게 되는 시점이 오게 되는데, 그 시점이 알고리즘 진행이 끝난 시점이며 군집화 작업이 종료된 후 나뉜 군집들을 실제 노드들의 군집 개수로 사용한다.

k-평균 알고리즘의 경우 먼저 군집의 수를 정하기 때문에 가장 좋은 군집상태를 만들기 위해서는 k를 다양한 수로 두고 군집화를 진행하여야 한다. 군집의 개수는 9개부터 12개까지의 경우를 살펴보았으며, 그 중 각 군집의 노드 개수가 고르게 분포되고 다양한 성격이 군집화 될 수 있는 경우가 11개라는 테스트 결과를 통해서, 11개의 감정어휘 유형 집단을 최종 군집 개수로 선정하게 되었다. 그리고 11개의 군집을 기반으로

노드 집단을 구분하기 위해 노드를 색상으로 구분했다. 그 결과 본 연구에서 제안한 히트맵 시각화 및 감정어휘의 유형에 따라 색상으로 구분된 최종 감정어휘 기반 영화 네트워크의 모습은 그림 7과 같이 나타낼 수 있다. 하나의 노드는 각각 영화 하나를 나타내며, 각각의 영화가 가지고 있는 감정어휘 빈도에 따라서 노드가 위치하게 된다. 그리고 노드 하나에는 각 영화의 자세한 감정어휘 빈도 및 정보를 보여줄 수 있도록 영화의 제목, 영화 포스터, 히트맵 시각화를 제공한다.

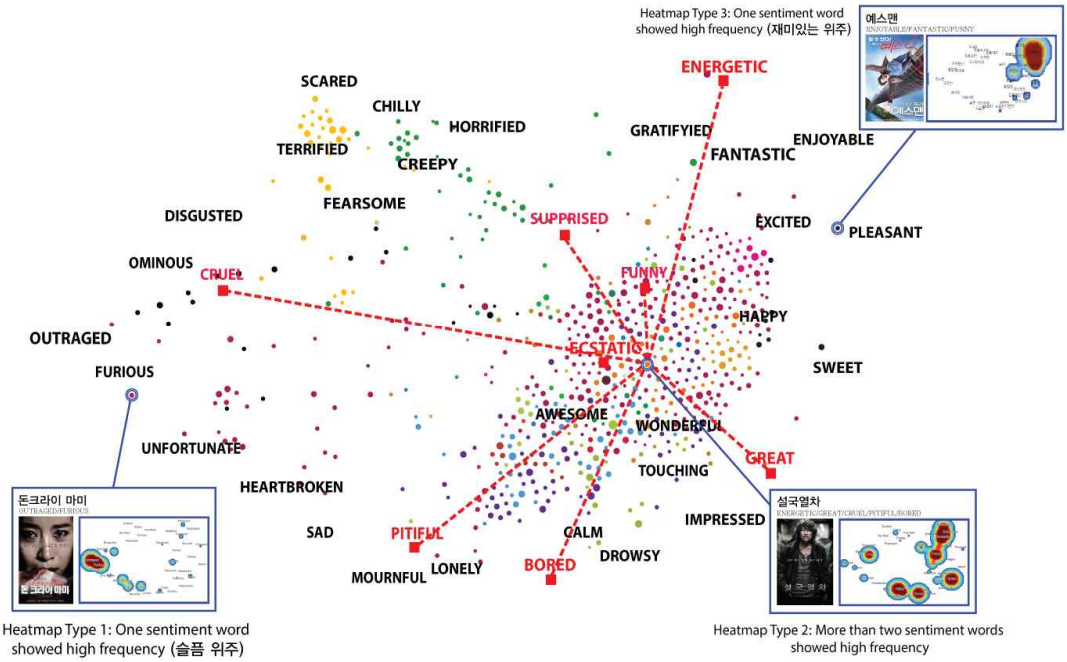

Fig. 7. Sentiment Movie Network (678 Movie nodes) & Heatmap visualization

3. Evaluation

3.1 Purpose and method of usability measurement experiment

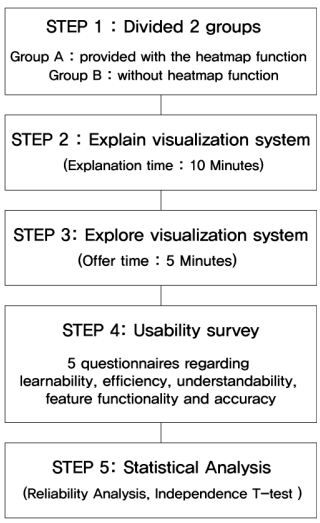

Fig. 8. Evaluation Process

본 연구에서 개발한 시각화의 검증을 위해 히트맵 시각화 기능을 제공하는 집단과 제공하지 않는 집단에 대해 시각화 사용에 대해 사용자가 느끼는 사용 용이성에 차이가 있는지를 확인하는 목적으로 실험을 설계했다. 히트맵 시각화의 유무에 따른 차이로 가설을 정한 이유는 히트맵 시각화가 네트워크를 이루고 있는 노드들이 가지는 감정어휘 정보를 가장 잘 나타낼 수 있는 수단이므로 사용 용이성 차이에 큰 영향을 미칠 수 있다고 생각했기 때문이다.

실험은 각각의 사용 용이성에 대한 세부 문항으로 구성된 설문지 실험으로 척도는 리커트 7점 척도를 사용하였다. 실험 대상은 각 집단 별로 시각화 분야에 대한 지식을 지니고 현재 데이터 시각화 분야를 공부중인 대학교 학생들을 표본으로 설정하였으며 총 100명을 대상으로 실험을 하였다. 이 중에서도 데이터가 누락되거나 설문 문항에 성실히 응답하지 않은 40부를 제외하고 60부의 설문 데이터를 최종 자료로 사용하였다. 검증 실험의 가설 및 검증 체계를 정리하면 아래의 내용 및 그림 8과 같다.

귀무가설 : 히트맵 시각화 기능 제공에 따른 두 집단 사이에는 유효한 차이가 없다.

대립가설 : 히트맵 시각화 기능 제공에 따른 두 집단 사이에는 유효한 차이가 있다.

3.2 Reliability analysis

데이터를 분석하기에 앞서 측정도구 의 신뢰성을 검증하기 위해 신뢰도 분석(Reliability Analysis)을 실시하였다. 신뢰도 분석은 문항 간 내적 일관성(Internal Consistency)을 측정하

는 방법으로 크론바하 알파(Cronbach's α)계수를 사용하여 이를 파악한다. 크론바하 알파 계수는 0에서 1의 값을 가지며 1에 가까울수록 문항의 신뢰도가 높다고 할 수 있다. 일반적으로 0.6이상의 값을 가지면 신뢰성이 있다고 하며 개별항목들을 하나의 척도로 종합하여 분석하는 것이 특징이라고 할 수 있다. 최종 측정 데이터를 활용하여 신뢰도를 분석하면 Table 2와 같다.

Table 2. Result of the Reliability Analysis.

Categories	Statements	Cronbach's α (Provide) & (Non-Provide)	Total
Learnability	1. It is easy to select a movie based on the sentiment words.	.698 (Provide) .827 (Non-Provide)	.666
Efficiency	2. It is efficient to select the node based on the sentiment of the movie.	.698 (Provide) .826 (Non-Provide)	
Understandability	3. It is easy to understand the sentiment distribution depending on varying node locations.	.661 (Provide) .747 (Non-Provide)	
Feature Functionality	4. It provides an adequate function to help user choose a movie.	.663 (Provide) .749 (Non-Provide)	
Accuracy	5. The selected movie and the sentiment distribution predicted from the movie's map coordinate matches.	.742 (Provide) .725 (Non-Provide)	

Table 3. Result of the Reliability Analysis ('a' group = Heatmap, 'b' group = No heatmap).

Question	P-value	Equal Variance Assumption	T-value	P-value	P-value / Alternative Hypothesis Adoption
1_a * 1_b	0.08203	Heteroscedasticity	4.8295	0.00003**	Adopt**
2_a * 2_b	0.5064	Heteroscedasticity	7.2038	0.00000001**	Adopt**
3_a * 3_b	0.2327	Heteroscedasticity	4.7609	0.000032**	Adopt**
4_a * 4_b	0.07771	Heteroscedasticity	4.3814	0.00011**	Adopt**
5_a * 5_b	0.0026	Equal variance	4.9205	0.000036**	Adopt*

신뢰도 분석결과 두 집단에 대한 각 항목을 제거할 때 크론바하 알파(Cronbach's α)계수의 최댓값은 히트맵 시각화를 제공한 경우, 정확성 문항을 제거 시 가장 높았고(0.742) 히트맵 시각화를 제공하지 않은 경우, 학습 용이성 문항을 제거 시 가장 높았다(0.827). 또한, 모든 항목에 대한 크론바하 알파 값이 0.6이상이므로 전체 문항에 대한 내적 일관성이 높고 따라서 신뢰도가 높다고 할 수 있다.

3.3 Average comparison per group

평균 비교는 독립표본 T검정(Independence T-test)을 실시하였는데 독립표본 T검정이란 두 집단이 각각 $N(\mu_{1,}\sigma_1^2)$과 $N(\mu_{2,}\sigma_2^2)$인 정규분포를 따르고 서로 독립이라는 가정 하에 두 집단 간 모평균에 차이가 있는지를 검정한다. 본 연구에서는 데이터의 정제 과정을 통해 60부의 설문 데이터를 두 집단으로 나누어 각각 30부의 설문 데이터를 사용하였으며 이는 중심 극한 정리에 의해 정규 분포를 가정하며 실험에 참여한 두 집단이 독립임을 가정하고 평균 비교를 시행하였다. 두 그래프에 대해 집단을 나누고 실험을 한 결과를 평균비교로 분석한 결과는 Table 3과 같다.

히트맵 시각화 기능을 제공한 집단과 히트맵 시각화 기능을 제공하지 않은 집단에 대해 평균비교 분석을 실시한 결과, 모든 문항에서 유의한 차이가 있다는 것을 확인하였다. Table 4는 대립가설을 채택한 문항에 대한 세부 사항이다.

귀무가설을 기각하고 대립가설을 채택한 모든 문항들을 세부적으로 확인한 결과, 히트맵 시각화 기능을 제공한 집단의 문항 평균 수치가 높게 나온 것을 확인할 수 있다. 이를 통해 시각화에서 히트맵 시각화 기능을 제공 할 때와 제공하지 않을 때, 모든 요인 문항에서 두 집단 사이간의 유효한 차이가 있다고 해석 할 수 있다.

Table 4. Details on the Statements with Alternative Hypothesis.

Question	95% confidence	Provide	Non-Provide
1_a * 1_b	1.1286< μ <2.7714	5.45	3.5
2_a * 2_b	1.8688< μ <3.3312	5.95	3.35
3_a * 3_b	1.1188< μ <2.7812	5.8	3.85
4_a * 4_b	0.9908< μ <2.7092	5.65	3.8
5_a * 5_b	1.2541< μ <3.0459	5.75	3.6

IV. Movie Recommendation System Scenario

이 장에서는 본 연구의 네트워크 시각화가 실제 영화추천의 기능으로서 어떻게 활용되는지 가상의 상황을 전제로 시나리오를 제시한다. 또한, 네트워크 시각화를 이용하여 영화추천을 제공하는 기존 사례인 'Popcha'[17]를 기반으로 본 연구의 영화추천시스템이 Popcha와 비교했을 때 어느 점에서 우수성을 나타내는 지 서술하고자 한다. 우선, 시나리오의 자세한 내용은 아래와 같다.

시나리오: 사용자가 본 영화를 기반으로, 감정이 비슷한 군집에서 영화 추천 받기

본 시나리오에서는 사용자가 관람한 영화의 감정어휘를 기반으로, 비슷한 감정을 갖는 영화 군집에서 영화를 추천 받는 상황을 전제로 한다. 이 시나리오에서 사용자가 관람한 영화는 'Star Track Into Darkness 2013'이며, 'Star Track'을 중심으로 영화추천을 받고자 한다. 1) 우선 사용자는 자신이 관람한 'Star Track' 노드를 네트워크에서 탐색하여 찾은 다음 해당 노드가 있는 곳을 선택하게 된다. 2) 다음으로 'Star Track' 주변에 있는 영화 노드가 무엇이 있는지 선택하게 된다. 3) 선택된 영화 노드는 그림 9와 같이 A, B 2개의 노드라 가정한다.

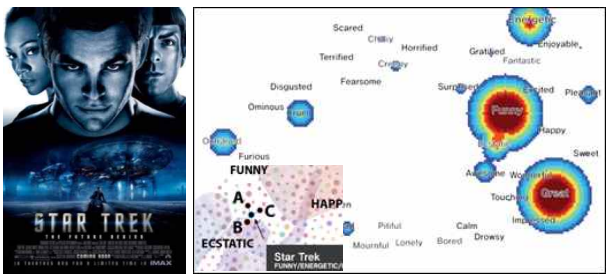

Fig. 9. Movie 'Star trek' heatmap visualization & discover 2 movies A, B which are located closely

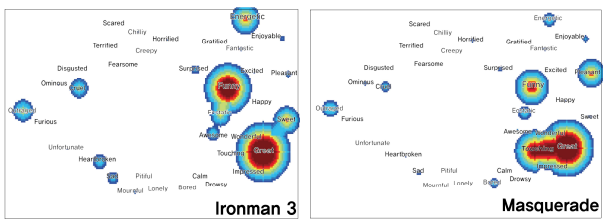

Fig. 10. 'Ironman 3'& 'Masquerade' heatmap visualization

4) 노드 A는 '아이언맨 3(Ironman 3)'라는 영화인데, 이 영화는 그림 10처럼 'Star Track'과 가장 유사한 히트맵 감정어휘 분포를 보이고 있기 때문에 이 사용자가 유사한 감정을 느끼고 싶다면 '아이언맨 3(Ironman 3)'를 보면 된다. 노드 B는 '광해: 왕이 된 남자(Masquerade)'라는 영화인데, 이 영화는 Star Track과 비슷한 히트맵 시각화 감정분포를 가짐과 동시에 '감동적인(Touching)' 부분이 추가되었기 때문에 이 사용자가 Star Track과 비슷하지만 감동적인 감정을 더 원한다면 '광해: 왕이 된 남자(Masquerade)' 영화를 보면 된다.

이처럼 사용자들은 자신이 경험한 영화를 바탕으로 유사한 영화를 추천 받고 싶을 때, 네트워크의 구조를 파악하여 해당 영화와 상대적으로 가까이에 있는 노드들을 후보로 정하고 각각의 후보 노드들의 감정어휘 빈도를 히트맵 시각화를 통해 분석한다면, 보다 효율적으로 보고 싶은 영화를 찾을 수 있을 것이라 예상한다.

다음으로, 네트워크 기반 영화추천 사례 중 하나인 'Popcha'(그림 11)와 본 연구의 시각화를 비교 분석하여 두 시스템이 어떤 차이가 있는 지, 그리고 본 연구에서 제안한 영화추천시스템이 가지는 장점이 무엇인지를 언급하고자 한다.

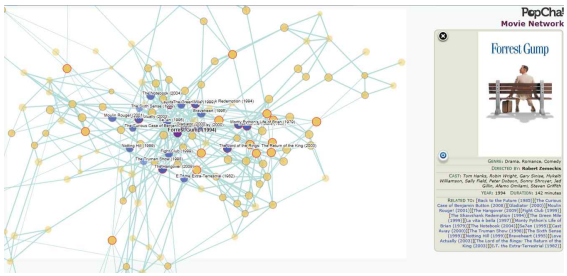

Fig. 11. 'Popcha' Movie Network visualization

Popcha는 영화에 대한 정보 중, '감독, 출연배우, 개봉연도, 장르, 스토리'의 정보를 기반으로 유사도를 계산하여 영화를 추천해주는 네트워크 시스템이다[17]. 따라서 사용자가 네트워크의 노드를 하나 선택했을 때, 그와 관련된 영화정보를 '감독, 출연배우, 개봉연도, 장르, 스토리'에 기반을 두어 제시한다. 그러나 이러한 시스템은 사용자가 선택한 영화에 대해서 감독, 출연배우, 스토리 등에 대한 사전지식이 없을 때는 추천결과를 공감하기 어려운 상황이 올 수 있으며, 자신이 관람한 영화를 기반으로 추천을 받기 위해서는 네트워크의 노드를 일일이 탐색해야하는 불편함이 있다.

반면에 본 연구의 추천 시스템은 '영화를 봤을 때 느낄 수 있는 감정어휘'를 기반으로 네트워크를 구성하였기 때문에 사용자가 영화 하나를 선택했을 때, 그 영화에 대한 사전지식이 없다 하더라도 2차원 감정어휘 분포맵과 히트맵 시각화를 통해서 감정 정보를 쉽게 인지할 수 있는 장점이 있다. 또한, 자신이 본 영화를 네트워크에서 찾고자 할 때에도 그 영화를 봤을 때 느낀 감정을 기억하고 있다면 쉽게 찾을 수 있다. 이러한 점을 미루어 볼 때, 본 연구의 시스템은 감정을 기반으로 다수의 사용자가 공감할 수 있는 영화 추천 경험을 제공하는 데에 있어 기존의 연구사례에 비해 효과적임을 알 수 있다.

V. Conclusion

본 연구에서는 집단 지성의 영화리뷰 데이터를 다차원 감정어 네트워크 시각화로 표현한 뒤 이를 직관적이고 의미적으로 해석하기 위한 세 가지 방법론을 제시하였다. 첫 번째는 개별 노드의 감정어휘 정보를 나타내는 히트맵 시각화(Heatmap Visualization)를 제공하였으며, 두 번째는 2차원 감정어휘 분포맵을 기준으로 네트워크 노드가 표현되는 방법을 제시했다.

본 연구의 후반부에는 고안된 방법들을 검증하기 위한 실험을 시행하였다. 그 결과 대부분의 사용자들이 노드의 위치와 히트맵에 관계에 대해 비교적 잘 인지한다는 것을 알 수 있었다. 또한, 각 노드를 이해할 때 히트맵 시각화가 적용된다면, 감정 정보 전달의 용이성이 향상되기 때문에 사용자들이 각 개별 영화가 갖고 있는 감정어휘를 이해하는 데 큰 도움을 준다는 것을 알게 되었다. 검증과정의 결과에 따라 본 연구의 두 가지 방법론에 대한 효과를 정리하면 다음과 같다. 우선 히트맵 시각화는 서로 비슷한 위치에 있는 노드에 대해서 감정어휘 분포의 미세한 차이점을 보여줄 때 적합하다는 것을 알 수 있다. 감정어휘 기반 영화 네트워크(Sentiment-Movie Network)는 영화가 가지고 있는 감정어휘 빈도 정보에 따라 노드(node)를 배치하기 때문에 네트워크를 해석할 때 해당 노드의 성격을 빨리 파악한다는 장점이 있다. 또한, 감정어휘 정보를 보여주는 2차원 분포맵은 감정어휘군의 군집에 따라서 영화 노드의 대표 감정을 파악하는 데 도움을 준다는 것을 볼 수 있다. 사용성 검증 실험이외에도 본 연구에서는 네트워크 시각화 및 히트맵 등을 활용하여 영화를 추천 받는 과정을 시나리오 형태로 제시함으로써, 사용자들이 감정에 따라 보고 싶은 영화를 효율적으로 선택할 수 있음 시사했다. 그리고 기존의 영화추천시스템 사례를 분석하여 비교함으로써, 본 연구에서 제안한 영화추천시스템이 가지는 장점이 무엇인지를 언급하였다. 이는 곧 본 연구의 방법론이 영화를 추천하는 데에 있어 사용자들에게 기존과는 다른 스타일의 추천을 할 수 있다는 가능성을 보여주었다.

본 연구는 향후 감정 분석(sentiment analysis)뿐만 아니라 온톨로지 구조 데이터에 대해서도 분석 작업을 시행할 예정이며, 온톨로지 구조 데이터가 가지고 있는 다양한 기준 및 의미 전달력을 향상시킬 수 있는 다차원 감정어 시각화를 만드는 것을 새로운 목표로써 채택할 예정이다.

REFERENCES

[1] M.J. Thomas, M. Edward, "Graph Drawing by Force-direct Placement," Software-practice and experience, Vol. 21, pp. 1129-1164, Nov. 1991.

[2] M. Kim, J. Kim, M. Cha, S. Chae, "An Emotion Scanning System on Text Documents," Korean Journal of the science of Emotion, Vol. 12, No. 4, pp. 433-442, Dec. 2009.

[3] Y. Lee, J. Jeong, "A Study on the Analysis of Emotion-expressing Vocabulary for Realtime Conversion of Avatar's Countenances," Korean Society of Design Science, Vol. 17, No. 2, pp. 199-208, May. 2004.

[4] J. Seong, K. Cho, "The Perceived Lexical Space for Haptic Adjective based on Visual Texture aroused form Need for Touch," Society of Design Convergence, Vol. 38, pp. 117-128, Feb 2013.

[5] D.W. Oard, M. Gary, "A conceptual framework for text filtering process," Software-practice and experience, Master's Thesis of Maryland University, 1998.

[6] B. Sarwar, G. Karypis, J. Konstan, J. Riedl, "Item-based Collaborative Filtering Recommendation Algorithms," Proceeding of the 10th International World Wide Web, pp. 285-295, May 2001.

[7] P. Li, S. Yamada, "A Movie Recommender System Based on Inductive Learning," Proceeding of IEEE Conference, Cybernetics and Intelligent Systems, pp. 318-323, Dec. 2004.

[8] C, Dunne, B, Shneiderrman, "Motif simplification: improving network visualization readability with fan, connector, and clique glyphs", Proceeding of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13), pp.3247-3256, May. 2013.

[9] N. Henry, A. Benzerianos, J. Fekete, "Improving the Readability of Clustered Social Networks using Node Duplication", IEEE Transactions on Visualization and Computer Graphics, Vol. 14, No. 6, pp. 1317-1324, Dec. 2008.

[10] D. Hahn., H. Kang, "Appropriateness and Frequency of Emotion Terms in Korea". Korean Journal of Psychology: General, Vol. 19, No. 2, pp.78-98, June. 2000.

[11] NAVER Movie, *http://movie.naver.com*

[12] Mecab-ko-analyzer, *http://eunjeon.blogspot.kr*

[13] L. Wilkinson,, M. Frendly, "The History of the Cluster Heat Map", The American Statistician, Vol. 63, No. 2, pp. 179-184, Sep. 2009.

[14] G. Robert, G. Nick, K. Rose, S. Emre, S. Awali, C,

Dunne, B, Shneiderrman, “Meirav Taieb-Maimon NetVisia: Heat Map, Matrix Visualization of Dynamic Social Network Statistics&Content”, Proceeding of Privacy, Security, Risk and Trust(PASSAT) and 2011 IEEE Third International Conference on Social Computing(SocialCom), pp. 19-26, 2011.

[15] Y. Jeong, Y. Chung, J. Park, “Visualisation of efficiency coverage and energy consumption of sensors in wireless sensor networks using heat map”, IET Communications, Vol. 5, No. 8, pp. 1129-1137, Sep. 2010.

[16] H, Ha, G, Kim, K. Lee, “A Study on Analysis of Sentiment Words in Movie Reviews and the Situation of Watching Movies”, Society of Design Convergence, Vol. 43, pp. 17-32, Dec. 2013.

[17] Popcha, http://bl.ocks.org/paulovn/9686202.

Authors

Hyo Ji Ha received the B.S. degrees in Digital Media from Ajou University, Korea, in 2013. And progress on M.S and Ph.D degrees in Life media Interdisciplinary Program, at Ajou University, Suwon, Korea, in 2013.

He is joined the graudate school of Lifemedia interdisciplinary program at Ajou University, Suwon, Korea, in 2013. He is currently Ph.d course in the Life media Interdisciplinary Program, Ajou University. He is interested in Information visualization, and Visual Analytics, User Experience Design.

Hyun Woo Han received the B.S. degrees in Digital Media from Ajou University, Korea, in 2014.

M.S. joined Life media Interdisciplinary Program, at Ajou University, Suwon, Korea, in 2014.

He is currently a Ph.d course in the Life media Interdisciplinary Program, Ajou University. He is interested in information visualization.

Seong Min Mun received the B.S degrees in Bachelor of Science from Pyeongtaek University and received the M.S degrees in Media Content from Ajou University, Korea, in 2014 and 2016, respectively. He is in Ph.D dual degree program between Ajou University for the Media Content and Paris 10 University for the Langage de Science. He is interested in text mining, opinion mining and data visualization.

Sung Yun Bae received the B.S. degrees in Digital Media from Ajou University, Korea, in 2015.

M.S. joined Life media Interdisciplinary Program, at Ajou University, Suwon, Korea, in 2015.

She is currently a master degree in the Life media Interdisciplinary Program, Ajou University. She is interested in information visualization.

Ji Hye Lee is in progress on M.S. degrees in Lifemedia interdisciplinary program from Ajou University, Korea. respectively. She is joined the graudate school of Lifemedia interdisciplinary program at Ajou University, Suwon, Korea, in 2015. She is currently a student research in the Department of Digital Media department, Ajou University. She is interested in User experience evaluation, Information visualization, and qualitative data analysis.

Kyung Won Lee received the MFA degree in computer graphics and interactive media from the Pratt Institute, USA, in 2002.

He joined the faculty of the Department of Digital Media at Ajou University, Suwon, Korea, in 2003.

He is currently a Professor in the Department of Digital Media at Ajou University. His research interests include information visualization, human-computer interaction, and media art.