

추천을 위한 키워드 가중치를 이용한 멀티모달 미디어 콘텐츠 분류

강지수¹, 백지원¹, 정경용^{2*}

¹경기대학교 컴퓨터과학과 학생, ²경기대학교 컴퓨터공학부 교수

Multimodal Media Content Classification using Keyword Weighting for Recommendation

Ji-Soo Kang¹, Ji-Won Baek¹, Kyungyong Chung^{2*}

¹Student, Department of Computer Science, Kyonggi University, South Korea

²Professor, Division of Computer Science and Engineering, Kyonggi University, South Korea

요 약 모바일 시장의 확장과 함께 멀티모달 미디어 콘텐츠의 제공을 위한 플랫폼이 다양해지고 있다. 멀티모달 미디어 콘텐츠에는 이종데이터들이 복합적으로 포함되어 있어 사용자가 선호 콘텐츠를 선택하기 위해 시간과 노력이 요구된다. 따라서 본 논문에서는 추천을 위한 키워드 가중치를 이용한 멀티모달 미디어 콘텐츠 분류를 제안한다. 제안하는 방법은 멀티모달 미디어 콘텐츠의 텍스트 데이터에서 키워드 가중치를 통해 콘텐츠를 가장 잘 나타내는 키워드를 추출한다. 추출된 키워드를 기반으로 서브클래스를 갖는 장르 클래스를 생성하고 이에 적절한 멀티모달 미디어 콘텐츠를 분류한다. 또한 개인화된 추천을 위해 사용자의 선호도 평가를 진행하여 사용자의 콘텐츠 선호도 분석 결과를 기반으로 멀티모달 콘텐츠를 추천한다. 성능평가는 추천 결과의 정확도와 만족도를 통해 우수함을 검증한다. 이는 사용자가 선호하는 장르와 키워드를 모두 고려하여 추천하기 때문에 정확도는 74.62%, 만족도는 69.1%로 높게 나타난다.

주제어 : TF-IDF, 텍스트 마이닝, 데이터마이닝, 분류, 멀티모달 콘텐츠, 추천

Abstract As the mobile market expands, a variety of platforms are available to provide multimodal media content. Multimodal media content contains heterogeneous data, accordingly, user requires much time and effort to select preferred content. Therefore, in this paper we propose multimodal media content classification using keyword weighting for recommendation. The proposed method extracts keyword that best represent contents through keyword weighting in text data of multimodal media contents. Based on the extracted data, genre class with subclass are generated and classify appropriate multimodal media contents. In addition, the user's preference evaluation is performed for personalized recommendation, and multimodal content is recommended based on the result of the user's content preference analysis. The performance evaluation verifies that it is superiority of recommendation results through the accuracy and satisfaction. The recommendation accuracy is 74.62% and the satisfaction rate is 69.1%, because it is recommended considering the user's favorite the keyword as well as the genre.

Key Words : TF-IDF, Text Mining, Data Mining, Classification, Multimodal Content, Recommendation

*This work was supported by the GRR program of Gyeonggi province [2017-B04, Image/Network-based Intellectual Information Manufacturing Service Research.

*Corresponding Author : Kyungyong Chung(dragonhci@gmail.com)

Received February 22, 2019

Revised March 27, 2019

Accepted May 20, 2019

Published May 28, 2019

1. 서론

스마트폰의 보급과 이동통신 기술의 발달에 따라 모바일 환경에서 멀티모달 미디어 콘텐츠를 제공하는 오픈 미디어 플랫폼이 등장하고 있다[1]. 이에 따라 사용자들은 방대한 양의 멀티모달 콘텐츠에 노출된다. 멀티모달 콘텐츠는 [이미지+텍스트], [음성+이미지] 등의 두 개 이상의 데이터 형태로 구성된다. 이는 직관적으로 내용을 이해하기는 편리한 장점이 있지만 주관적인 데이터가 포함되는 문제가 있다. 따라서 다른 콘텐츠들보다 선택에 시간이 보다 더 오래 소요되며 만족스러운 결과를 얻기 힘들다. 이러한 문제점을 해결하기 위해 추천시스템이 지속적으로 연구되고 있다[2]. 또한 효율적으로 사용자에게 콘텐츠를 추천하기 위해 콘텐츠 간의 상관관계를 파악한 후 여러 속성을 반영하고 이를 필터링하여 콘텐츠 기반 필터링을 통한 추천 방법이 연구되고 있다[3]. 콘텐츠 기반 필터링은 콘텐츠의 특성을 분석하여 콘텐츠 자체 또는 콘텐츠와 사용자의 선호도 유사성을 활용하여 추천하는 방법이다[4]. 이는 사용자의 과거 기록을 통해 유사한 콘텐츠를 추천하는 방법이다. 이때 이전에 이용한 콘텐츠와 비슷한 콘텐츠만 추천되는 과도한 특수화(Over Specialization) 문제점이 발생한다[5]. 이러한 문제를 보완하기 위해 본 논문은 추천을 위한 키워드 가중치를 이용한 멀티모달 미디어 콘텐츠 분류를 제안한다. 제안하는 방법은 TF-IDF를 기반으로 개발한 TF-IoCF(Term Frequency - Inverse of Content Frequency)를 이용해 아이탬을 대표하는 키워드를 추출한다. TF-IDF는 문서의 특징 추출에 많이 사용되는 방법이다. 특정 문서 내에서 어떤 단어가 많이 나왔을 때 해당 문서 군을 모두 검사하여 그 단어가 모든 문서군내에 흔하게 나오는 단어인지, 특정 문서에서만 나오는 독특한 단어인지 판단한다[6]. 흔하게 나오는 단어일 경우 단어의 중요도를 낮추고 반대의 경우에는 중요도를 높여 가중치를 조절하는 기법이다. 따라서 개인화된 멀티모달 콘텐츠 추천을 위해 멀티모달 콘텐츠를 적절한 클래스에 분류한다. 분류된 클래스의 특성과 사용자의 선호도를 이용하여 보다 정확한 멀티모달 미디어 콘텐츠 추천이 가능하다.

2. 관련연구

2.1 텍스트 분석을 위한 자연어 처리

텍스트 마이닝은 반정형 및 비정형 텍스트 데이터에서 자연어처리와 문서처리 기술을 이용한다. 이를 통해 방대한

텍스트 데이터에서 유용한 정보를 추출할 수 있다[7]. 이는 문서군집, 문서분류, 문서요약, 정보추출 등의 기술 분야를 다루며 자료 처리과정과 분석과정으로 구분된다. 처리과정은 비정형 데이터를 정제하고 가공하여 분석에 용이한 형태로 만드는 과정이다. 분석과정은 머신러닝이나 데이터 마이닝을 통해 정제된 텍스트 데이터로부터 유의미한 정보를 찾아내는 과정이다[8].

자연어 처리 기술은 인간이 발화하는 자연어를 컴퓨터가 이해할 수 있도록 변환하거나 컴퓨터가 처리한 결과를 인간이 이해할 수 있는 언어로 표현한다. 형태소 분석은 문장을 구성하는 단어로부터 최소 의미단위인 형태소를 분리해 각 형태소의 문법적 기능에 따라 적절한 품사를 부착하거나 단어의 원형을 복원하는 기술이다[9]. 형태소 분석은 음절 기반의 계산과 기계학습 접근법을 이용한다. 음절 기반 계산은 형태소 후보의 발생과 선택의 과정으로 나누고 형태소 후보의 수를 급격히 줄여 사전에 접근하는 액세스를 감소시킴으로써 모델의 성능을 높인다. 특징 함수를 통해 음절의 특징을 형식화해 파악하여 사전 확인 전에 부적절한 형태소 후보를 제거한다[10]. 기계학습 접근법은 응집성이 높고 복합적인 언어의 형태소 분석을 한다. 이는 시퀀스 라벨링과 커널 방법 훈련을 기반으로 자연어의 형태소 특징의 비선형 관계를 다방면으로 추출한다. 이를 통해 형태학적으로 응집성이 강한 자연어에 대한 형태소 분석의 정확성을 향상시킨다[11].

3. 추천을 위한 키워드 가중치를 이용한

멀티모달 미디어 콘텐츠 분류

3.1 데이터 전처리 및 속성변수 선별

본 논문에서는 다양한 멀티모달 콘텐츠 중 영화를 선정하여 추천을 진행한다. 대부분의 영화 아이탬은 제작사 또는 보급사에서 제공하는 짧은 글의 줄거리 데이터를 갖는다. 멀티모달 콘텐츠는 해당 콘텐츠를 표현할 수 있는 적절한 키워드를 파악하기 어렵다[12]. 따라서 줄거리를 텍스트 마이닝하여 키워드를 도출한다. Gropulens[13]에서 제공하는 총 27,278개의 아이탬의 목록을 수집한다. 또한 TMDb(The Movie Database)[14]에서 크롤링을 통해 아이탬과 관련된 데이터를 준비한다. 수집된 데이터의 속성은 아이탬의 제목, 장르, 줄거리이다. 이러한 속성들 중 줄거리에서 텍스트 마이닝을 통해 분류를 위한 키워드를 선별한다. 이는 N-gram방법을 이용해 주어진 아이탬의 줄거리에 포

함된 단어들을 추출한다. N-gram은 N개의 단위의 창(window)을 만들어 왼쪽에서 오른쪽으로 이동하며 해당 gram의 빈도수를 기록하여 문자열의 특징을 추출하는 방법이다 [15]. 이를 2개의 단어 단위(biword)로 설정하여 줄거리를 분석하고, 빈도순으로 정렬한다. 이때 특정한 단어가 줄거리 내에 얼마나 자주 등장하는지를 나타내는 값인 TF[16]를 사용하여 빈도수가 높은 상위 3개의 단어를 키워드로 선택한다. 하지만 줄거리에서 특정 단어의 빈도가 높게 나오는 모든 경우가 중요한 키워드라고 판단하기 힘들다. 만약 많은 아이템 내용의 줄거리에서 빈번하게 등장하는 단어일 경우 그 키워드는 아이템 내에서 공통적으로 많이 등장하는 키워드로 중요도가 높지 않다. 따라서 문서군내에서 등장하는 키워드에 가중치를 두어 중요도를 높여주는 TF-IDF 기반 TF-IoCF를 제안한다. 제안하는 TF-IoCF(Term Frequency - Inverse of Content Frequency)는 기존 TF-IDF를 멀티모달 미디어 콘텐츠에 적합하게 개량한 것으로 TF와 IoCF를 곱하여 구한다. IoCF는 전체 콘텐츠 중에서 한 콘텐츠가 나타나는 빈도 값이며 콘텐츠의 전체 개수에서 단어를 포함한 콘텐츠 요약의 수를 나눈 값에 로그를 취한 값이다. IoCF값이 클수록 다른 콘텐츠에서 잘 등장하지 않고 해당 콘텐츠에만 등장하는 독특한 단어이다. 때문에 TF-IoCF를 통해 아이템을 가장 잘 표현할 수 있는 키워드 선택이 가능하다.

Table 1. Movie Keyword Extracted through Text Mining

Title	Genre	Keyword	
		Word	TF-IoCF
The Thieves	Action, Crime, Drama	Thief	0.03283
		Plan	0.01691
		Diamond	0.01216
Extreme Job	Comedy	Chicken	0.03075
		Drug	0.02844
		Investigation	0.01363
Iron Man	SF, Action	Suit	0.03555
		Army	0.01582
		Weapon	0.01068
...

Table 1은 텍스트 마이닝을 통해 추출한 아이템 키워드를 나타낸다. 데이터는 아이템의 제목, 최소 1개에서 최대 3개의 장르, 3개 키워드의 속성을 갖는다.

3.2 멀티모달 미디어 콘텐츠의 클래스 분류

멀티모달 미디어 콘텐츠 추천을 위해 클래스 분류 기법을 사용한다. 이를 위해 멀티모달 콘텐츠가 분류될 클래스

를 미리 정의한다. 클래스를 정의하기 위해서는 전처리된 총 2만 7천여 개의 아이템 데이터를 학습 데이터로 사용하고 1차적으로 장르를 기준으로 장르 클래스를 정의한다. 이때 {Comedy}, {Comedy, Romance}, {Comedy, Drama, Romance}와 같이 동일한 장르를 포함하더라도 모두 다른 세 가지의 클래스로 처리한다. 장르의 속성은 총 14개로 조합을 통해 총 527개의 장르 클래스로 분류된다. 장르 클래스에 포함되는 키워드 서브클래스는 각 장르 클래스마다 최소 5개, 최대 15개의 키워드를 대표로 하여 구성한다. 구성된 서브클래스에 해당 키워드를 포함하고 있는 아이템들을 분류한다. 만약 키워드 서브클래스의 대표 키워드를 2개 이상 포함하고 있는 아이템의 경우에는 TF-IoCF값이 더 높은 키워드의 서브클래스로 분류한다. Fig. 1은 {Thriller, Action} 클래스가 포함하고 있는 서브클래스를 나타낸다. {Thriller, Action} 클래스는 Zombie, Spy, Race 등의 키워드로 서브클래스로 구성된다. 아이템 줄거리의 텍스트 마이닝을 통해 Zombie를 대표 키워드로 가진 아이템은 부산행, 월드 워 Z, 레지던트 이블 등이 있다. 이와 같은 일련의 과정으로 장르 클래스, 키워드 서브클래스를 정의하고 27,278개의 아이템을 적절한 클래스에 분류한다.

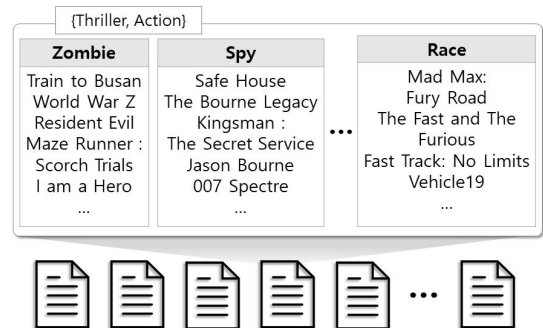


Fig. 1. Sub-classes included {Thriller, Action} Class

3.3 분류 클래스 기반의 멀티모달 미디어 콘텐츠 추천

멀티모달 미디어 콘텐츠 추천을 위해 새로운 사용자의 초기 아이템의 선호도를 평가한다. 이는 사용자가 선호하는 장르 선택, 그에 따른 세부키워드 입력의 순서로 진행된다. 먼저 총 14개의 단일 장르 속성 중에서 사용자가 선호하는 장르를 선택한다. 예를 들어, 사용자가 {Animation}, {Musical}의 장르를 선택했다면, 두 개의 장르 클래스뿐만 아니라 이들을 모두 포함하는 {Animation, Musical} 클래스 또한 선호할 가능성이 높다. 이와 같이 선택된 장르들의 조합을 고려하여 사용자가 선호하는 장르의 클래스를 파악한다.

다음으로 사용자가 선호하는 아이템의 키워드를 입력한다. 사용자가 선택한 장르 클래스에서 입력받은 키워드를 탐색하여 서브클래스 별로 후보목록집합을 생성한다. 생성된 후보목록집합의 키워드 서브클래스에서 TF-IoCF 값을 비교하여 값이 가장 높은 아이템들을 선별한다. 이와 같은 방법으로 사용자의 선호 장르 클래스와 키워드의 초기평가가 완료되면 이를 기반으로 개인화된 아이템 콘텐츠를 추천한다. 만약 사용자가 선호하는 장르로 {Animation}, {Musical}을 선택하고 키워드로 {가족}을 입력할 경우 각 장르와 그 조합에서 해당 키워드 서브클래스를 찾아 검색하여 추천한다. 이때에는 {Animation, Musical} 장르의 '겨울왕국'과 {Musical} 장르의 '맘마미아!', {Animation} 장르의 '니모를 찾아서'가 추천된다. 두 개의 키워드를 입력할 경우 장르별로 두 개씩 총 6개의 아이템이 추천된다. 또한 협력적 필터링을 사용하기 위해 사용자와 비슷한 클래스를 선호하는 사용자들의 아이템 선호도를 이용한다. 사용자와 유사성을 가진 다른 사용자들을 선별하여 사용자와 겹치지 않는 또 다른 클래스의 아이템을 추천한다. 이 과정에서 사용자의 선호 클래스를 더욱 확장시킴으로써 비슷한 콘텐츠만 지속적으로 추천되는 과도한 특수화를 방지한다.

Fig. 2는 제안하는 추천을 위한 키워드 가중치를 이용한 멀티모달 미디어 콘텐츠 분류의 프로세스를 나타낸다. 첫 번째 단계에서 멀티모달 콘텐츠의 데이터를 수집하여 데이터 전처리과정을 통해 속성 변수를 선별한다. 두 번째 단계는 해당 멀티모달 콘텐츠를 적절한 장르 클래스와 키워드 서브클래스에 분류하는 과정이다. 마지막으로 추천을 위해 사용자가 선호하는 장르와 키워드를 평가받는다. 이를 분석하여 두 번째 단계에서 분류한 장르와 키워드 서브클래스를 이용하여 멀티모달 미디어 콘텐츠를 추천한다. 따라서 다양하고 정확한 추천이 가능하다.

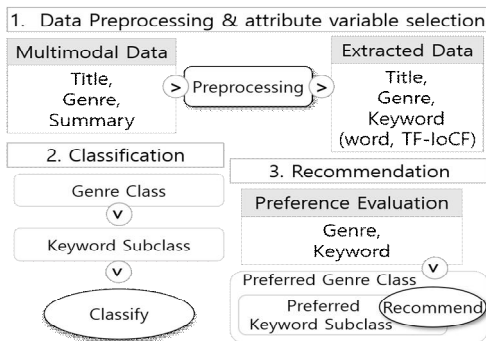


Fig. 2. Process of Multimodal Media Content Classification using Keyword Weighting for Recommendation

4. 성능 평가

제안하는 키워드 가중치를 이용한 추천 방법(KWCR, Keyword Weighting Classification Recommendation)의 성능을 평가한다. 제안하는 방법과 TF-IDF를 이용한 키워드 기반 추천(TIKR, TF-IDF Keyword Recommendation)[17], 그래프 유사도 기반 추천(GSR, Graph Similarity Recommendation)[18]의 정확도와 만족도를 비교한다. 평가 데이터는 온라인과 오프라인 커뮤니티에서 200명을 대상으로 선호 장르, 키워드, 최근 만족스럽게 본 10개의 영화와 별점, 가장 감명 깊게 본 영화 5개와 별점을 수집한 데이터를 활용한다. 추천의 정확도 평가는 사용자의 선호 장르와 키워드를 입력하여 제안하는 KWCR 방법과 기존의 TIKR와 GSR 방법으로 영화를 추천하고 사용자의 실제 선호 여부를 판단한다. 추천 시스템으로 추천된 아이템이 실제 사용자가 응답한 최근 만족스럽게 본 10개의 영화와 가장 감명 깊게 본 영화 5개의 포함 여부와 이 중에서도 높은 평점의 영화를 추천했는지 평가한다. 예를 들어, 사용자 A의 최근 만족스럽게 본 영화 {m1, m2, m3, ..., m10}, 가장 감명 깊게 본 영화 {m11, m12, ..., m15}와 이에 대한 각각의 평점이 있고, 사용자가 선호하는 장르 {j1, j2}, 키워드 {k1, k2}가 있다. 이에 기반을 둔 사용자 선호도에 따라 추천된 영화 {r1, r2, r3, ..., rn}까지 n개의 영화를 비교하여 정확한 추천이 되었는지 평가한다. 이를 통해 추천된 아이템들 중에서 사용자가 실제로 좋아하는, 기호에 만족되는 아이템이 추천되었는지에 대해 평가한다. 추천 정확도는 선호하는 영화를 맞춘 정도를 나타내며 식(1)과 같다.

$$Accuracy = \frac{1}{200 \times 15} \times \sum_{i=1}^{200} P_i$$

(1)

P_i 는 사용자의 선호에 맞는 아이템을 맞춘 개수이며 이에 대한 평균에 입력받은 선호 영화 개수인 15를 나누어 소수 다섯째자리에서 반올림한다. 또한 추천에 대한 만족도는 추천된 영화에 대한 사용자의 만족도의 지표로써 식(2)와 식(3)과 같다.

$$S_k(\%) = \frac{1}{N \times 5} \times \sum_{i=1}^N IRS \times 100 \quad (2)$$

$$Satisfaction(\%) = \frac{1}{200} \times \sum_{k=1}^{200} S_k \times 100 \quad (3)$$

IRS(Item Rating Score)는 사전에 사용자가 평가한 선호하는 아이템에 대한 평가 점수이다. 사용자는 1부터 5까지의 정수로 선호도에 가중치를 주어 평가하며 숫자가 클수록 선호도가 높음을 의미한다[19,20]. 식(2)는 개인 사용자(k)의 추천 만족도 S_k 를 구하는 식으로 추천 방법이 추천한 영화 중 자신이 선호하는 아이템 개수 N 개의 IRS 값을 모두 더 한 뒤 $N \times 5$ 로 나눈 값의 평균으로 구한다. 이를 이용해 식(3)에서 총 200명의 사용자 추천 만족도(S_k)의 평균을 구한다.

Table 2. Performance Results of Recommendation Method

Recommend method	Accuracy of Recommendation (%)	Satisfaction with Recommendation (%)
TIKR	71.17	36.3
GSR	49.61	64.4
KWCR	74.62	69.1

Table 2는 추천 방법에 따른 성능평가 결과를 나타낸다. 성능평가 결과 제안하는 방법을 이용한 KWCR 추천 방법이 TIKR, GSR 방법 보다 각각 약 3.45%p, 25.01%p 만큼 더 높은 정확도를 보이며, 추천된 영화에 대한 사용자의 전체적인 만족도가 69.1%로 다른 모델보다 32.8%p, 4.7%p 더 높게 나타난다. TIKR를 이용한 추천 방법은 사용자가 제공하는 키워드만을 이용하여 추천하는 방법으로 키워드에 대한 정확도는 71.17%로 높게 나타나지만, 사용자의 취향과 전혀 다른 장르의 영화가 추천되는 경우가 발생하여 전체적인 만족도 평가가 36.3%로 낮게 나타난다. GSR를 이용한 추천 방법은 영화간의 그래프 유사도를 기반으로 하여, 동일한 시리즈가 연속으로 추천되는 문제로 전체적인 추천 정확도가 49.61%로 낮게 나타난다. 결과적으로 제안하는 추천을 위한 키워드 가중치를 이용한 멀티모달 미디어 콘텐츠 분류는 장르뿐만 아니라 키워드도 고려하여 선호도를 산출하기 때문에 보다 더 높은 추천 정확도와 추천 만족도를 갖는다.

5. 결론

본 논문에서 제안하는 추천을 위한 키워드 가중치를 이용한 멀티모달 미디어 콘텐츠 분류는 텍스트 마이닝을 통하여 멀티모달 콘텐츠의 키워드를 추출한다. 추출한 키워드를 이용해 멀티모달 콘텐츠를 분류함으로써 분류를 세밀하게 하여 추천 시스템의 성능을 높인다. 또한 분류 클래스를 이용하여 사용자의 선호도를 수집하고 분석하여 선호클래스를 선별한다. 선별된 선호클래스를 통해 사용자에게 개인화된 멀티모달 미디어 콘텐츠를 추천할 수 있으며 협력적 필터링을 사용함으로써 보다 다양한 추천이 가능하다. 추천에 대한 정확도는 74.62%로 높게 나타나고 추천 콘텐츠에 대한 만족도 또한 69.1%로 우수하다.

다양한 멀티모달 콘텐츠에 제안하는 방법을 적용하여 멀티모달 추천 시스템을 구축할 수 있다. 이를 통해 개인화된 멀티모달 콘텐츠 추천 서비스를 제공할 수 있다.

REFERENCES

- [1] J. Park, M. Kim & S. Rho. (2015). A Study for Personalized Multimedia Information Services. *The Journal of Society for e-Business Studies*, 20(3), 79–87.
- [2] J. H. Kim, D. Lee & K. Chung. (2014). Recommendation based on Context-aware Model for Personalized u-Healthcare Service. *Multimedia Tools and Applications*, 71(2), 855–872.
- [3] J. Son & S. B. Kim. (2017). Content-based filtering for recommendation systems using multiattribute networks. *Expert Systems with Applications*, 89, 404–412.
- [4] Y. Wu & A. L. P. Chen. (2000). Index structures of user profiles for efficient Web page filtering services. *Proceedings 20th IEEE International Conference on Distributed Computing Systems*, 644–651. DOI : 10.1109/ICDCS.2000.840981
- [5] M. Balabanović & Y. Shoham. (1997). Fab : content-based, collaborative recommendation. *Communications of the ACM*, 40(3), 66–72.
- [6] A. Aizawa. (2003). An information-theoretic perspective of tf-idf measures. *Information Processing & Management*, 39(1), 45–65.
- [7] S. Choi, J. Jeon, B. Subrata & O. Kwon. (2015). An Efficient Estimation of Place Brand Image Power Based on Text Mining Technology. *Journal of Intelligence and Information Systems*, 21(2), 113–129.
- [8] A. Hotho, A. Nürnberger & G. Paass. (2005). A Brief Survey of Text Mining. *LDV Forum*, 20, 19–62.

- [9] T. Ritchey. (2006). Problem structuring using computer-aided morphological analysis. *Journal of the Operational Research Society*, 57(7), 792-801.
- [10] S. S. Kang & Y. T. Kim. (1994). Syllable-based model for the Korean morphology. *Association for Computational Linguistics*, 1(15), 221-226.
- [11] V. Dhanalakshmi, M. Anandkumar, R. U. Rekha, C. Arunkumar, K. P. Soman & S. Rajendran. (2009). Morphological Analyzer for Agglutinative Languages Using Machine Learning Approaches. *International Conference on Advances in Recent Technologies in Communication and Computing*, 433-435.
DOI : 10.1109/ARTCom.2009.184
- [12] J. Park, M. Kim & S. Rho. (2015). A Study for Personalized Multimedia Information Services. *The Journal of Society for e-Business Studies*, 20(3), 79-87.
- [13] GroupLens. (2016). *MovieLens*. <https://grouplens.org/>
- [14] TMDb. (2016). *The Movie Database*. <https://www.themoviedb.org/>
- [15] P. F. Brown, P. V. Desouza, R. L. Mercer, V. J. D. Pietra & J. C. Lai. (1992). Class-based n-gram models of natural language. *Computational linguistics*, 18(4), 467-479.
- [16] G. Salton & C. Buckley. (1988). Term-weighting approaches in automatic text retrieval. *Information processing & management*, 24(5), 513-523.
- [17] S. Thomaidou & M. Vazirgiannis. (2011). Multiword keyword recommendation system for online advertising. *IEEE*, 423-427.
DOI : 10.1109/ASONAM.2011.70
- [18] C. W. Hang & M. P. Singh. (2010). Trust-based recommendation based on graph similarity. *In Proceedings of the 13th International Workshop on Trust in Agent Societies*, 82.
- [19] J. C. Kim & K. Chung. (2019). Prediction Model of User Physical Activity using Data Characteristics-based Long Short-term Memory Recurrent Neural Networks. *KSPII Transactions on Internet and Information Systems*, 13(4), 2060-2077.
- [20] J. C. Kim & K. Chung. (2019) Mining based Time-Series Sleeping Pattern Analysis for Life Big-data. *Wireless Personal Communications*, 105(2), 475-489.

강지수(Kang, Ji-Soo)

[학생회원]



- 2016년 2월 ~ 현재 : 경기대학교 컴퓨터과학과 학부생
- 2017년 2월 ~ 현재 : 경기대학교 컴퓨터과학과 데이터마이닝 연구실 연구원
- 관심분야 : 데이터마이닝, 헬스케어, 빅데이터 분석, 추천 시스템
- E-Mail : kangjs920@gmail.com

백지원(Baek, Ji-Won)

[학생회원]



- 2013년 3월 ~ 2017년 2월 : 상지대학교 컴퓨터정보공학부 (공학사)
- 2017년 2월 ~ 2018년 3월 : (주)인피닉스 연구원
- 2018년 9월 ~ 현재 : 경기대학교 컴퓨터과학과 (석사과정)

- 2018년 9월 ~ 현재 : 경기대학교 컴퓨터과학과 데이터마이닝 연구실 연구원
- 관심분야 : 데이터마이닝, 딥러닝, 빅데이터 분석, 헬스케어, 인공지능
- E-Mail : hy1233hh@naver.com

정경용(Chung, Kyungyong)

[정회원]



- 2000년 2월 : 인하대학교 전자계산공학과 (공학사)
- 2002년 2월 : 인하대학교 전자계산공학과 (공학석사)
- 2005년 8월 : 인하대학교 컴퓨터정보공학부 (공학박사)

- 2006년 3월 ~ 2017년 2월 : 상지대학교 컴퓨터정보공학부 교수
- 2017년 3월 ~ 현재 : 경기대학교 컴퓨터공학부 교수
- 관심분야 : 데이터 마이닝, 헬스케어, 빅데이터, 지능시스템, 인공지능, HCI, 정보검색, 추천 시스템
- E-Mail : dragonhci@hanmail.net