

사용자 선호도와 군집 알고리즘을 이용한 퍼지-계층적 분석 기법 기반 영화 추천 시스템

오재택*, 이상용**

공주대학교 컴퓨터공학과*, 공주대학교 컴퓨터공학부**

A Movie Recommendation System based on Fuzzy-AHP with User Preference and Partition Algorithm

Jae-Taek Oh*, Sang-Yong Lee**

Dept. of Computer Science & Engineering, Kongju National University*

Div. of Computer Science & Engineering, Kongju National University**

요 약 현재 추천 시스템은 실제 사용자가 선호하는 항목을 추천하는지, 아니면 단순히 관심 정도의 항목을 추천하는지 알 수 없다는 문제와 사용자들이 매우 적어 적합한 항목을 추천할 수 없는 데이터 희소성 문제, 새로운 사용자들이 유입됨에 따라 사용자들이 만족하는 항목을 추천하기 위해 시스템의 성능이 저하되는 Cold-Start 문제 등이 발생한다. 본 연구에서는 이러한 문제점을 해결하기 위해 불확실한 상황이나 문제들을 반영할 수 있는 퍼지-계층적 분석(Fuzzy-Analytic Hierarchy Process)과 주어진 항목들을 비슷한 항목들끼리 모으는 데이터 군집화 알고리즘을 활용하여 사용자들에게 만족할 수 있는 영화를 추천하기 위한 시스템을 구현하였다. 61명을 대상으로 영화 선호도에 대한 설문 조사를 실시한 데이터를 본 시스템에 적용한 결과 Fuzzy-AHP 기법을 통해서 데이터 희소성 문제를 해소할 수 있었으며, 또한 데이터 군집화 알고리즘을 통해 새로운 사용자들이 유입되어도 사용자에게 적합한 항목이 추천되었음을 확인할 수 있었다. 향후 노이즈 데이터나 아웃라이어(Outlier) 데이터를 걸러낼 수 있는 밀도 기반 클러스터링에 대한 연구가 필요할 것으로 생각된다.

주제어 : 데이터 희소성, 콜드-스타트, 퍼지-계층적 분석 기법, 데이터 군집화 알고리즘, 추천 시스템

Abstract The current recommendation systems have problems including the difficulty of figuring out whether they recommend items that actual users have preference for or have simple interest in, the scarcity of data to recommend proper items due to the extremely small number of users, and the cold-start issue of the dropping system performance to recommend items that can satisfy users according to the influx of new users. In an effort to solve these problems, this study implemented a movie recommendation system to ensure user satisfaction by using the Fuzzy-Analytic Hierarchy Process, which can reflect uncertain situations and problems, and the data partition algorithm to group similar items among the given ones. The data of a survey on movie preference with 61 users was applied to the system, and the results show that it solved the data scarcity problem based on the Fuzzy-AHP and recommended items fit for a user with the data partition algorithm even with the influx of new users. It is thought that research on the density-based clustering will be needed to filter out future noise data or outlier data.

Key Words : Data Sparsity, Cold-Start, Fuzzy-Analytic Hierarchy Process, Data Partition Algorithm, Recommendation System

Received 28 September 2017, Revised 30 October 2017
Accepted 20 November 2017, Published 28 November 2017
Corresponding Author: Sang-Yong Lee
(Kongju National University)
Email: sylee@kongju.ac.kr

ISSN: 1738-1916

© The Society of Digital Policy & Management. All rights reserved. This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. 서론

추천 시스템이란 사용자의 정보를 바탕으로 선호도를 예측하여 적합한 항목을 제공하는 시스템을 말한다[3, 9].

이러한 추천 시스템은 실제 사용자가 선호하는 항목을 추천하는지, 아니면 단순히 관심 정도의 항목을 추천하는지 알 수 없는 문제와 더불어 전체 항목 수에 비해 사용자가 이용한 항목 수가 적어 사용자에게 적합한 항목을 추천할 수 없는 상황을 말하는 데이터 희소성(Data Sparsity) 문제, 그리고 새로운 사용자들이 지속적으로 유입됨에 따라 새로운 사용자들이 만족하는 항목을 추천하기 위해 시스템의 성능이 저하되는 문제를 말하는 Cold-Start 문제 등이 발생하게 된다[6, 11, 12, 17].

이러한 문제점으로 사용자에게 적합한 항목을 올바르게 추천할 수 없어 사용자는 추천 시스템에 의존하기 보다 선호하는 항목의 평점을 활용하거나 직접 이용하는 방법을 선택하게 된다[10].

최근 추천 시스템의 연구 동향으로는 행렬 분해(Matrix Factorization) 기법과 데이터 임퓨테이션(Data Imputation) 기법을 사용하여 Cold-Start 문제와 데이터 희소성 문제를 해소하고 있다. 행렬 분해 기법은 주어진 데이터 정보 행렬을 항목(Item)의 정보나 사용자의 정보 등의 행렬로 분해하여 잠재적인 요소들을 추출한 다음, 다시 조합하여 추천 시스템의 성능을 높이는 기법으로 주로 Cold-Start 문제를 해소하기 위해 적용되어 왔다. 그러나 계산 복잡도가 비교적 높다는 단점이 있다[6, 13].

데이터 임퓨테이션 기법은 사용자가 평가하지 않은 항목(Item)에 대한 선호도를 예측하는 기법으로 주로 데이터 희소성 문제를 해소하기 위해 적용되어 왔다. 그러나 이 기법은 MovieLens와 같은 데이터 세트(Data Set)를 이용하여 제안된 기법으로 실제 데이터를 수집하고 적용하였을 때 추천 시스템의 성능이 높아졌는지 검증이 되지 않았다는 문제가 있다[12].

본 연구에서는 위와 같은 문제점들을 해소하기 위해 계산 복잡도가 비교적 복잡하지 않으며, 사용자의 주관적인 선호도를 직접적으로 반영할 수 있는 Fuzzy-AHP 기법과 데이터 군집화 알고리즘인 K-means 클러스터링을 이용하여 사용자에게 적합한 항목을 추천하고자 하였다.

Fuzzy-AHP는 불확실한 상황이나 다양한 항목들에

대한 선호도를 순위화하여 우선순위가 높은 항목을 추천하는 문제해결 형 의사결정 방법론으로 실제 사용자의 수가 극히 적어도 주관적인 선호도를 객관적으로 알 수 있어 데이터 희소성 문제를 해소하는데 가장 적합하다고 할 수 있으며, 또한 데이터 군집화 알고리즘은 주어진 항목들을 비슷한 항목들끼리 모으는 작업을 말하는 알고리즘으로 새로운 사용자들이 지속적으로 유입되어도 기존의 선호도 데이터들에 대한 군집화를 통해 군집 추천이 가능해지기 때문에 Cold-Start 문제를 해소하는데 가장 적합하다고 할 수 있다[2].

이러한 관점에서 본 연구에서는 Fuzzy-AHP 기법을 적용하여 사용자의 주관적인 선호도를 직접적으로 반영하고, K-means 클러스터링을 이용하여 사용자들의 선호도 군집화를 통해 보다 적합한 항목을 추천하기 위한 영화 추천 시스템을 제안하고자 한다.

2. 관련연구

2.1 Fuzzy-AHP

Fuzzy-AHP는 Da-Yong Chang 등이 제안한 방법으로 삼각 퍼지 수를 이용하여 척도를 언어적인 표현방식으로 표현하여 나타나는 모호성과 애매함 등의 문제를 보완하여, 보다 정확한 의사결정을 도와주는 의사결정 방법론이다[7, 11, 14, 16].

Fuzzy-AHP는 척도에 대한 언어적인 표현방식을 명확한 수치로 나타내는 것보다 간격을 주어 표현할 수 있는 삼각 퍼지 수를 사용하여 AHP(Analytic Hierarchy Process)의 문제점을 보완하였다.

삼각 퍼지 수(Triangular Fuzzy Numbers)는 퍼지 수의 소속 함수(Membership Function) 형태 중에서도 가장 많이 사용되고 있는 삼각형 꼴의 소속 함수이다. 삼각 퍼지 수는 어떤 수치에 대하여 하한 값(l), 중앙 값(m), 상한 값(u)으로 구성된다. 기본적으로 삼각 퍼지 수의 소속 함수는 연구자의 기본적인 주관에 따라 결정되며, 표준적인 소속 함수는 존재하지 않는다[8].

본 연구에서 사용된 삼각 퍼지 수의 소속 함수는 <Table 1>와 같으며, Fuzzy-AHP의 의사결정에서 널리 사용되고 있다[11].

<Table 1> Triangular Fuzzy Numbers for Pairwise-comparison

Scale	Linguistic Scale	Triangular Fuzzy Numbers	Triangular Fuzzy Reciprocal Numbers
1	Equal Preference	(1, 1, 2)	(1/2, 1, 1)
2	Median Preference	(1, 2, 3)	(1/3, 1/2, 1)
3	Moderate Preference	(2, 3, 4)	(1/4, 1/3, 1/2)
4	Median Preference	(3, 4, 5)	(1/5, 1/4, 1/3)
5	Strong Preference	(4, 5, 6)	(1/6, 1/5, 1/4)
6	Median Preference	(5, 6, 7)	(1/7, 1/6, 1/5)
7	Very Strong Preference	(6, 7, 8)	(1/8, 1/7, 1/6)
8	Median Preference	(7, 8, 9)	(1/9, 1/8, 1/7)
9	Extreme Preference	(8, 9, 9)	(1/9, 1/9, 1/8)

Fuzzy-AHP를 사용하여 의사결정을 진행하기 위해서는 총 8단계로 나뉘는데, 처음 2단계와 마지막 2단계는 AHP의 의사결정 방식의 진행을 따른다[7, 10, 11, 14, 16].

- 1단계:** 평가하고자 하는 주제와 관련된 모든 기준들과 항목들을 검토하여 계층 구조를 구성한다.
- 2단계:** 평가 항목들과 세부 평가 항목들 간의 비교를 통해 계층별로 쌍대비교 행렬을 작성한다.
- 3단계:** 작성한 쌍대비교 행렬을 삼각 퍼지 수로 변환된 쌍대비교 행렬로 다시 작성한다.
- 4단계:** 삼각 퍼지 수로 변환된 쌍대비교 행렬을 이용하여 퍼지 합성 확장 값(Value of Fuzzy Synthetic Extent, S_i)을 계산한다.

변환된 쌍대비교 행렬 B 의 원소 a_{ij} 의 속성 i 번째의 퍼지 합성 확장 값은 식(1)과 같이 계산한다.

$$S_i = \left(\sum_{j=1}^n l_{ij}, \sum_{j=1}^n m_{ij}, \sum_{j=1}^n u_{ij} \right) \otimes \left(\sum_{i=1}^n \sum_{j=1}^n l_{ij}, \sum_{i=1}^n \sum_{j=1}^n m_{ij}, \sum_{i=1}^n \sum_{j=1}^n u_{ij} \right)^{-1} \quad (1)$$

- 5단계:** 퍼지 합성 확장 값을 이용하여 가능성 정도(Degree of Possibility, d')를 계산한다.

가능성 정도는 식(2)와 같이 계산한다.

$$d'(S_i) = \min V(S_i \geq S_j) \quad (2)$$

$$(i = 1, 2, \dots, n, i \neq j)$$

또한 식(2)는 식(3)을 이용하여 계산한다.

$$V(S_1 \geq S_2) = \begin{cases} 1, & \text{if } m_1 \geq m_2 \\ \frac{l_2 - u_1}{(m_1 - u_1) - (m_2 - l_2)} & \text{else if } u_1 \geq l_2 \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

- 6단계:** 가능성 정도를 이용하여 상대적 중요도의 합이 1이 되도록 정규화를 한다.
- 7단계:** 계층별로 구한 상대적 중요도를 이용하여 상위 계층과 그와 관련된 하위 계층끼리 서로 곱하여 평가 항목별로 종합 중요도를 계산한다.
- 8단계:** 평가 항목별로 구한 종합 중요도를 이용하여 우선순위를 구하고, 우선순위가 높은 최적의 평가 항목을 선택한다.

본 연구에서는 Fuzzy-AHP 기법을 이용하여 사용자의 주관적인 선호도를 직접적으로 반영하고, 반영된 선호도를 이용하여 사용자들에게 적합한 영화를 추천할 수 있도록 하였다.

2.2 K-means

K-means 클러스터링은 기하학적 클러스터링에 가장 널리 사용되는 비지도 학습 기법 중 하나의 알고리즘으로써 1965년 Forgy와 1967년 McQueen에 의해 제안되었다. 종종 Lloyd 알고리즘으로 알려져 있다[15].

K-means 클러스터링은 유클리드 공간(Euclidean Space)에 n 개의 데이터로 구성된 데이터 세트 D 에 대하여 K 개의 클러스터(Cluster)로 군집화하는 알고리즘을 말한다. 여기서 클러스터란 군집화된 K 개의 부분집합 중 하나를 말한다[4].

K 개의 클러스터로 군집화하기 위해서는 군집의 개수인 K 를 결정하고, 군집의 개수만큼 각 군집의 중심점이 되는 값을 임의로 정한다. 그리고 n 개의 데이터들을 임의로 정한 중심점을 기준으로 가장 가까운 군집에 할당한다. 각 데이터들과 중심점과의 거리는 유클리드 거리(Euclidean Distance)를 이용하여 식(4)와 같이 계산한다

[1, 4].

$n \times 1$ 벡터 X 와 벡터 $W_j (j=1, 2, \dots, k, \dots, n)$ 간의 유클리드 거리는 다음과 같다[5].

$$d = |X - W_j| = \left[\sum_{i=1}^n (x_i - w_{ik})^2 \right]^{\frac{1}{2}} \quad (4)$$

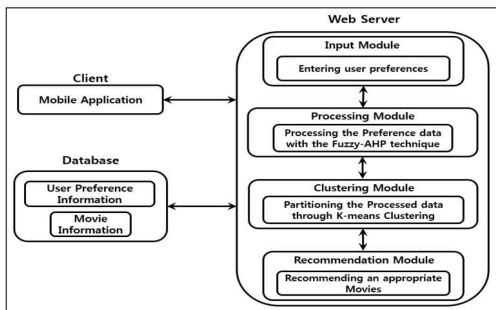
여기서 x_i 와 w_{ik} 는 각각 벡터 X 와 벡터 W_j 의 k 번째 원소이다.

그런 다음 할당된 데이터들을 중심으로 각 군집의 새로운 중심점을 다시 계산한다. 새로운 중심점이 임의로 정한 중심점과 차이가 있으면 각 군집에 할당된 데이터들을 새로운 중심점을 기준으로 가장 가까운 군집에 할당하고, 중심점과의 거리는 유클리드 거리를 이용하여 계산한다. 그리고 새로운 중심점이 임의로 정한 중심점과 차이가 없으면 n 개의 데이터들이 K 개의 클러스터로 군집화되었음을 최종적으로 알 수 있다[1, 4].

본 연구에서는 사용자들에게 보다 적합한 영화를 추천하기 위해 사용자들의 선호도를 K-means 클러스터링을 이용하여 군집화하였으며, 군집화를 통해 새로운 사용자들이 지속적으로 유입함에 따라 추천 시스템의 성능이 저하되는 Cold-Start 문제를 군집 추천을 이용하여 해소할 수 있도록 하였다.

3. 시스템 설계

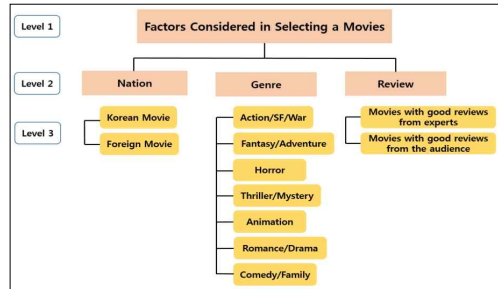
본 연구에서 제안하는 영화 추천 시스템은 [Fig. 1]과 같이 웹 서버와 데이터베이스 그리고 클라이언트로 구성된다.



[Fig. 1] System Structure

본 연구에서의 웹 서버는 Microsoft의 인터넷 정보 서비스를 사용하였으며, 각 모듈은 각각의 프로세스를 담당하게 된다.

우선, 입력 모듈은 사용자가 선호하는 영화에 대한 선호도를 입력하는 모듈로 [Fig. 2]과 같이 계층 구조로 구성된 AHP 계층도를 이용하여 의사결정을 진행하게 되며, 입력된 선호도는 데이터베이스에 저장하게 된다.



[Fig. 2] AHP Structure of Input Module

사용자가 선호하는 영화를 알기위해 구성된 계층 구조에서 각 평가 항목(Level 2)은 'Reddy'가 문헌 연구를 통해 제안한 영화 흥행이 성공적으로 이루어지기 위한 요인들의 내용을 참고하였으며, 각각의 세부 평가 항목 (Level 3)은 한국영화진흥위원회와 포털 사이트인 'Naver'에서 영화를 분류하는 기본적인 방법을 따랐다 [18, 19, 20].

처리 모듈은 데이터베이스에 저장된 사용자가 선호하는 영화의 선호도를 이용하여 Fuzzy-AHP 기법으로 처리하는 과정이다. 이 과정을 통해서 사용자가 선호하는 영화가 어떠한 분류의 영화인지 알 수 있다.

클러스터링 모듈은 처리 모듈에서 처리된 선호도 데이터를 이용하여 군집화하는 모듈이다. 이 모듈은 K-means 클러스터링을 이용하여 선호도 데이터를 군집화해 사용자가 선호하는 영화에 대해 전체적인 분포를 알 수 있어 추천 모듈을 통해 적합한 추천을 할 수 있도록 한다.

마지막으로 추천 모듈은 처리 모듈과 클러스터링 모듈을 거친 선호도 데이터를 이용하여 사용자가 선호하는 군집에 가까운 영화를 추천하는 모듈이다. 이 추천 모듈을 통해서 사용자가 선호하는 영화를 모바일 어플리케이션으로 확인할 수 있다.

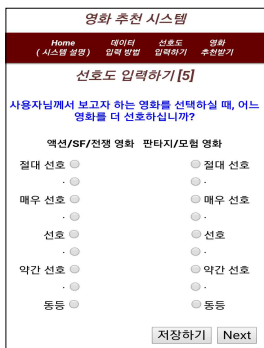
4. 시스템 구현

본 연구에서 제안하는 영화 추천 시스템을 구현하기 위한 개발 환경은 <Table 2>와 같다[10].

<Table 2> Development Environment

Mobile Device	Samsung Galaxy S8+
Operating System	Android 7.0 Nougat, Windows 7
Framework	ASP.NET 4.5.2
Language	Java for Android, XML, HTML5.0, CSS 3.0, Visual C#
Database	MS SQL Server 2012
Tool	Android Studio, Visual Studio 2015, R

모바일 장비로는 삼성전자의 갤럭시 S8+ 스마트폰을, 시스템 운영체제는 안드로이드 최신 버전인 'Nougat'와 'Windows 7'을 사용하였다. 웹 서버 개발 응용 소프트웨어로는 'ASP.NET 4.5.2'를 사용하여 웹 사이트를 구성하였으며, 프로그래밍 언어는 'Java for Android'와 'XML', 'HTML5.0', 'CSS 3.0' 그리고 'Visual C#'을 사용하여 'Android Studio'와 'Visual Studio 2015'를 통해 클라이언트를 구현하였다. 그리고 'R'을 이용하여 각 사용자 선호도 데이터를 군집화해 추천 시스템에 적용할 수 있도록 하였다. 마지막으로 데이터베이스는 'MS SQL Server 2012'를 사용하여 구축하였다[10].



[Fig. 3] [Program Capture] User Interface of Input Module

[Fig. 3]은 입력 모듈을 구현한 사용자 인터페이스로, 사용자가 선호하는 영화를 <Table 1>에 따른 9가지 척

도의 데이터를 입력받아 데이터베이스에 저장될 수 있도록 'ASP.NET'의 웹 폼으로 서버 컨트롤을 이용하여 웹 페이지를 구성하였고, 웹 페이지의 디자인을 구현하기 위해 HTML 5.0과 CSS 3.0을 사용하여 구현하였다.

[Fig. 4]은 처리 모듈과 클러스터링 모듈을 거친 선호도 데이터를 이용하여 사용자가 선호하는 군집에 가까운 영화를 추천하는 추천 모듈을 구현한 사용자 인터페이스로, 처리 모듈에서 계산된 종합 중요도가 가장 큰 순서로 우선순위가 계산되며, 우선순위가 가장 큰 영화에 대해 사용자들이 선호하는 군집에 가까운 영화를 추천한다. 우선순위를 계산하기 위한 알고리즘은 버블 정렬(Bubble Sort)을 사용하여 우선순위를 정렬하였다.



[Fig. 4] [Program Capture] User Interface of Recommendation Module

[Fig. 4]에서 개봉 중인 영화, 개봉되었던 영화를 대상으로 사용자가 가장 선호하는 장르인 스릴러/미스터리 (Thriller/Mystery Movie)를 추천하고 있으며, 각 영화 포스터를 터치하면 자세한 영화 정보를 볼 수 있다.

5. 실험 및 분석

5.1 실험

본 연구에서는 데이터 회소성과 Cold-Start 문제를 어떻게 해소하는지 확인하기 위하여 61명을 대상으로 영화 선호도에 대한 조사를 실시하였다. 설문 항목은 크게 두 가지로, 첫 번째는 영화 제작 국가, 영화에 대한 장르, 영화에 대한 평가의 쌍대비교, 두 번째는 평가 항목에 대한 각각의 세부 평가 항목으로 한국 영화와 외국 영화, 7가지의 장르, 전문가 평가가 좋은 영화와 관객 평가가 좋은

영화에 대한 쌍대비교의 설문을 조사하였다.

먼저 데이터 희소성 문제 해소를 확인하기 위하여 61명의 설문 데이터에서 60명의 데이터를 제외한 1명의 데이터만으로 실험을 하였다. 그리고 Cold-Start 문제 해소를 확인하기 위해서 60명의 설문 데이터를 기반으로 새로 유입된 1명의 설문 데이터를 사용하여 실험을 하였다.

5.2 분석

본 연구에서의 추천 시스템에서 새로 유입된 사용자가 선호하는 영화의 선호도를 Fuzzy-AHP로 처리한 결과는 <Table 3>와 같다.

<Table 3> The Preference of Movies Preferred by New User

Action/SF/War Movie	0
Fantasy/Adventure Movie	0.125
Horror Movie	0
Thriller/Mystery Movie	0.875
Animation	0
Romance/Drama Movie	0
Comedy/Family Movie	0

새로 유입된 사용자의 선호도는 <Table 3>와 같이 판타지/모험 영화(Fantasy/Adventure Movie)와 스릴러/미스터리 영화(Thriller/Mystery Movie)를 선호하는데, 그 중에서도 스릴러/미스터리 영화를 가장 선호하는 것으로 나타났다.



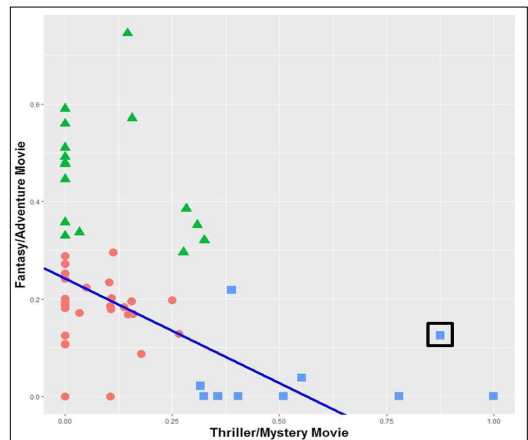
[Fig. 5] [Program Capture] The User Interface of a Recommendation by the Fuzzy-AHP Technique

<Table 3>의 선호도를 이용하여 데이터 희소성 문제 해소를 확인하기 위하여 60명의 데이터를 제외하고, 군집화 알고리즘을 포함하지 않은 Fuzzy-AHP 기법을 사용한 결과는 [Fig. 5]와 같다.

[Fig. 5]의 결과와 같이 추천 시스템을 사용하는 사용자가 부족해도 사용자에게 <Table 3>의 선호도 결과와 동일한 스릴러/미스터리 영화(브이아이피, 장산범, 더 잭팟, 몽타주, 프레스티지)와 2번째로 선호하는 장르인 판타지/모험 영화(혹성탈출, 발레리안, 스타더스트, 에이트 빌로우)가 추천된 것을 확인함으로써 데이터 희소성 문제가 Fuzzy-AHP 기법을 통해 해소됨을 알 수 있다.

또한 기존의 데이터 안에서 새로운 사용자가 유입되어도 사용자에게 적합한 항목을 추천하는지 알아보기 위해 <Table 3>의 선호도를 이용하여 군집화 알고리즘을 적용한 추천 시스템의 결과는 [Fig. 4]와 같다. [Fig. 5]의 결과와 달리 본 추천 시스템에서는 새로 유입된 사용자가 가장 선호하는 영화의 장르를 기준으로 모든 영화에 대한 군집화를 시작하게 되며, 군집화한 결과에서 새로 유입된 사용자의 선호도에 맞는 군집의 장르를 추천하게 된다.

[Fig. 6]은 'R'로 실행한 새로 유입된 사용자의 선호도 데이터와 60명의 선호도를 3개의 클러스터로 군집화한 스릴러/미스터리 영화와 판타지/모험 영화의 그래프이다. 최근 사용자의 선호도는 큰 '□'로 표시되어 있으며, 그래프에서 '○', '△', '□'는 3개의 클러스터로 군집화된 모습을 보여주고 있다.



[Fig. 6] The Graph of User Preferences Partitioning

[Fig. 6]을 살펴보면, 각각의 군집은 판타지/모험 영화를 선호하는 군집인 ‘△’ 분포와 판타지/모험 영화와 스릴러/미스터리 영화를 같이 선호하는 군집인 ‘○’ 분포와 스릴러/미스터리 영화를 선호하는 군집인 ‘□’ 분포로 군집화되어 있다. 그리고 회귀분석(Regression Analysis)을 통한 회귀선(Regression Straight Lines)이 감소함으로써 기존의 사용자들은 스릴러/미스터리 영화보다 판타지/모험 영화를 더 선호하는 것을 알 수 있다.

그러나 새로 유입된 사용자는 ‘□’ 군집에서 벗어난 아웃라이어(Outlier, 모든 클러스터에서 아주 멀리 떨어진 값)이지만, 스릴러/미스터리 영화를 선호하는 군집에 맞게 스릴러/미스터리 영화만을 추천한 결과를 [Fig. 4]에서 확인함으로써 Cold-Start 문제가 해소됨을 알 수 있다.

6. 결론

본 연구에서는 추천 시스템의 고질적인 문제인 데이터 희소성과 Cold-Start를 해결하기 위해 Fuzzy-AHP와 데이터 군집화 알고리즘인 K-means 클러스터링을 이용한 영화 추천 시스템을 구현하였다.

그 결과 Fuzzy-AHP 기법을 통해서 데이터 희소성 문제를 해소할 수 있음을 확인할 수 있었으며, 데이터 군집화 알고리즘을 통해 새로운 사용자들이 유입되어도 사용자에게 적합한 항목이 추천되었음을 알 수 있었다.

향후 대규모 데이터에서도 효과적으로 군집화할 수 있으며, 노이즈 데이터나 아웃라이어 데이터를 걸러낼 수 있는 밀도 기반 클러스터링(Density-based Clustering)에 대한 연구가 필요할 것으로 생각된다.

REFERENCES

- [1] Doo-Soon Park, Yang-Se Moon, Young-Hyo Park, Chan-Hyun Yoon, Young-Sik Jung, Hyung-Suk Jang, Co-Author, “Big Data Computing Techniques”, p.126-127, Hanbit Academy, Inc., 2015.
- [2] Hwa-Jong Kim, “Introduction to Data Science”, p.188, Hongreung Science Publishing Co., 2014.
- [3] Il Lim, “Recommendation System Using R”, pp.3-4, Chaosbook, 2015.
- [4] Jiawei Han, Micheline Kamber, Jian Pei, Co-Author, Sa-Bum Jung, Yong-Keun Song, Translation, “Data Mining: Concepts and Techniques”, pp.577-582, Acom, 2015.
- [5] Michael Negnevitsky, Author, Yong-Hyuk Kim, Translation, “Artificial Intelligence 2nd Ed”, p.269, Hanbit Academy, Inc., 2013.
- [6] Dong-Hyun Kim, Chan-Young Park, Gene-Oh Oh, Hwan-Jo Yu, “Research on Cold-Start Recommendation” Communications of KIISE, Vol. 34, No. 6, pp. 16-21, 2016.
- [7] Hyang-Soon Joun, Sang-Yong Lee, “Technical Entrepreneurship Education Service Quality Evaluation System based on FAHP”, Journal of Digital Convergence, Vol. 13, No. 10, pp. 509-516, 2015.
- [8] Hae-Chun Rhee, Kyu-Yong Lee, In-Jae Lee, “Fuzzy Set Theory As a Method of Policy Evaluation: With a Case of Job Creation Policy”, The Korea Association for Policy Studies, Vol. 16, No. 3, pp. 67-91, 2007.
- [9] Hong-Bok Lee, Dong-Ok Won, Seong-Whan Lee, “Collaborative Filtering Recommender System With Clustering and Histogram”, Proceedings of KISS Winter Conference 2016, pp. 588-590, 2016.
- [10] Jae-Taek Oh, Sang-Yong Lee, “AHP-Based Recommendation System of Mobile Games Reflecting User Preferences”, Journal of Digital Convergence, Vol. 15, No. 1, pp. 427-433, 2017.
- [11] Jae-Taek Oh, “Fuzzy-AHP Based Mobile Games Recommendation System Using Bayesian Network”, Dept. of Computer Science and Engineering the Graduate School of Kongju National University, 2017.
- [12] Jee-Woon Ha, Hyung-Wook Kim, Sang-Wook Kim, “Data Imputation Methods for Effective Collaborative Filtering”, Communications of KIISE, Vol. 34, No. 6, pp. 8-15, 2016.
- [13] Noo-Ri Kim, Han-Byul Bang, Bedeuro Kim, Sei-Hee Lee, Jee-Hyoung Lee, “Research Trends in Context-aware Recommender Systems”, Communications of KIISE, Vol. 34, No. 6, pp. 22-29, 2016.
- [14] You-Jin Park, “Analyzing the Efficiency of SCM Using Fuzzy-AHP/DEA”, Dept. of Business Administration

- the Graduate School of Yonsei University, 2013.
- [15] A. Vattani, "K-means Requires Exponentially Many Iterations Even in the Plane", Proceedings of the Twenty-Fifth Symposium on Computational Geometry 2009, pp. 324-332, 2009.
- [16] Da-Yong Chang, "Applications of the Extent Analysis Method on Fuzzy AHP", European Journal of Operational Research, Vol. 95, No. 3, pp. 649-655, 1996.
- [17] G. Adomavicius, A. Tuzhilin, "Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-art and Possible Extensions", IEEE Transaction on Knowledge Data Engineering, Vol. 17, No. 6, pp. 734-749, 2005.
- [18] S. K. Reddy, V. Swaminathan, C. M. Motley, "Exploring the Determinants of Broadway Show Success", Journal of Marketing Research, Vol. 35, No. 3, pp. 296-315, 1998.
- [19] <http://movie.naver.com/movie/sdb/rank/rmovie.nhn>
- [20] <http://www.kobis.or.kr/kobis/business/main/main.do>

오 재 택(Oh, Jae Taek)



- 2015년 2월 : 대전대학교 IT경영공학과 (공학사)
- 2017년 2월 : 공주대학교 컴퓨터공학과 (공학석사)
- 2017년 3월 ~ 현재 : 공주대학교 컴퓨터공학과 (박사 과정)
- 관심분야 : 인공지능, 추천 시스템
- E-Mail : ohjt15@kongju.ac.kr

이 상 용(Lee, Sang Yong)



- 1984년 2월 : 중앙대학교 전자계산학과 (공학사)
- 1988년 2월 : 일본동경공업대학대학원 총합이공학연구과 (공학석사)
- 1988년 3월 ~ 1989년 2월 : 일본 NEC 중앙연구소 연구원
- 1993년 2월 : 중앙대학교 일반대학원 전자계산학과 (공학박사)
- 1996년 9월 ~ 1997년 8월 : University of Central Florida 방문교수
- 1993년 8월 ~ 현재 : 공주대학교 컴퓨터공학부 교수
- 관심분야 : 인공지능, 컨텍스트 예측, 추천 시스템
- E-Mail : sylee@kongju.ac.kr