JKIICE Journal of the Korea Institute of Information and Communication Engineering

한국정보통신학회논문지(J. Korea Inst. Inf. Commun. Eng.) Vol. 19, No. 1 : 113~118 Jan. 2015

혼합 필터링 기반의 영화 추천 시스템에 관한 연구

정인용 · 양새동 · 정회경*

A Study on Movies Recommendation System of Hybrid Filtering-Based

In-Yong Jeong · Xitong Yang · Hoe-Kyung Jung*

Department of Computer Engineering, Paichai University, Daejeon 302-735, Korea

요 약

추천 시스템은 증가되고 있는 정보에서 사용자가 요구하는 적합한 정보를 선별해 제공해준다. 추천 시스템은 기존에 입력된 정보들을 알고리즘을 통해 선별하는 과정을 거치고 사용자의 정보나 내용 기반으로 정보를 제공한다. 추천 시스템의 문제점으로는 Cold-Start가 있으며, Cold-Start는 새로운 사용자의 정보가 충분하지 않아서 추천 시스템에서 새로운 사용자에게 정보를 추천할 때 발생한다. Cold-Start를 해결하기 위해선 사용자의 정보나 항목 정보가 충족해야 한다.

이에 본 논문에서는 협업 필터링 기법과 내용 기반의 필터링 기법을 혼합한 혼합 필터링 기법 기반으로 Cold-Start 문제를 해결하고 이를 사용하는 영화 추천 시스템을 제안한다.

ABSTRACT

Recommendation system is filtering for users require appropriate information from increasing information. Recommendation system is provides the information based on user information or content that information entered in the original through process of filtering through the algorithm. Recommend system is problems with Cold-start, and Cold-start is not enough information in the occurrences for new users of recommend system in the new information to the user when recommend. Cold-start is should meet to resolve the user of information and item information.

In this paper, Suggest for movie recommendation system on collaborative filtering techniques and content-based filtering techniques based to a hybrid of a hybrid filtering techniques to solve problems in cold-start

키워드 : Cold-Start, 내용 기반의 필터링 기법, 추천 시스템, 협업 필터링 기법, 혼합 필터링 기법

Key word : Cold-Start, Collaborative Filtering Technique, Content-Based Filtering Technique, Hybrid Filtering Technique, Recommend System

접수일자 : 2014. 11. 14 심사완료일자 : 2014. 12. 04 게재확정일자 : 2014. 12. 19

* **Corresponding Author** Hoe-Kyung Jung(E-mail:hkjung@pcu.ac.kr, Tel:+82-42-520-5640)
Department of Computer Engineering, Paichai University, Daejeon 302-735, Korea

Open Access http://dx.doi.org/10.6109/jkiice.2015.19.1.113 print ISSN: 2234-4772 online ISSN: 2288-4165

Ⅰ. 서 론

정보 기술의 발전으로 인해 정보의 생산은 급격하게 증가하고 있다. 이로 인해 사용자들은 찾고자 하는 정보를 쉽게 찾을 수 없는 현상이 발생되어 정보의 정확도가 저하되고 있다. 이러한 문제를 해결하기 위해 추천 시스템의 중요성이 높아지고 있다[1].

추천 시스템은 사용자의 취향이나 선호의 기반으로 정보를 선별하여 사용자에게 적합한 정보를 제공한다. 따라서 추천 시스템이 적용된 기업은 다른 기업에 비해 고객관리나 매출 향상 등의 경쟁력을 갖출 수 있으며 불필요한 정보를 제공하지 않아 자원의 낭비나 고객의 만족도를 높일 수 있다[2]. 추천 시스템에는 대표적으로 협업적 필터링과 내용 기반의 필터링을 사용한다[3-5]. 추천 시스템에서 사용되는 필터링은 사용자의 정보를 활용하여 적합한 정보를 제공하지만 새로운 사용자에게는 적합한 정보를 제공하지 못하는 Cold-Start[6] 문제가 발생한다.

본 논문에서는 협업 필터링 기법과 내용 기반의 필터링 기법을 혼합하여 Cold-Start 문제를 해결하고 추천 시스템에 영향이 될 수 있는 주요 알고리즘을 비교하여 효율적인 알고리즘을 분석하고 영화 추천 시스템을 제안한다.

Ⅱ. 관련연구

2.1. Cold Start

잠재적으로 발생되는 Cold-Start는 자동화된 데이터 모델에서 발생된다. 대표적으로 추천 시스템에서 발생하며 정보 선별에 기반이 되는 정보가 부족하면 발생된다. 협업 필터링 기법에서는 식별된 사용자가 선호하는 정보와 항목 기반으로 정보를 선별하지만 이러한 정보가 충분하지 않으면 Cold-Start가 발생된다. 그리고 내용 기반의 필터링 기법에서는 사용자의 정보나 상품의 평가 내용 기반으로 정보를 선별하지만 충분하지 않으면 Cold-Start가 발생된다.

2.2. 협업 필터링 기법

대부분의 추천 시스템에서는 협업 필터링 기법을 사용한다. 협업 필터링 기법을 사용한 추천 시스템은 사용자들의 선호도를 수집한 뒤 이를 기반으로 사용자들의 관심사나 유사한 취향을 예측한다. 협업 필터링 기법에서는 사용자 정보와 항목 기반으로 정보를 선별한다.

사용자 정보 기반의 협업 필터링[7]은 사용자들의 선호도나 상품의 평가들을 수집하고 비슷한 성향을 가진 사람들을 연결한다. 그리고 이를 기반으로 사용자에게 정보를 제공한다. 항목 기반의 협업 필터링 기법[8]은 사용자들이 평가한 상품들의 정보를 사용한다. 사용자들은 과거에 선호했던 제품들과 유사한 제품을 선호하는 경향이 있다는 점을 사용하여 유사한 제품들의 정보를 제공한다.

2.3. 내용 기반의 필터링 기법

내용 기반의 필터링 기법은 자연 언어 처리나 정보 검색 분야에 기반을 두고 있으며 정보의 내용이나 사용자의 정보들을 비교하여 사용자에게 적합한 정보를 제공해준다. 내용 기반의 필터링 기법에 사용되는 주요 모델은 불리언 모델과 벡터공간 모델, 확률모델 등과 같은 기법을 사용하며 사용자가 과거에 사용했거나 평가했던 상품의 유사도를 측정하여 정보를 제공한다.

Ⅲ. 시스템 설계

본 장에서는 Cold-Start 문제를 해결하기 위한 방안을 제안하고 웹 사이트 형식으로 구현했다. 제안하는 영화 추천 시스템의 구성은 세 단계로 구성되어 있다. 먼저 사용자 정보 수집 단계에서는 새로운 사용자의 정보를 시스템에 입력하고 데이터베이스에 등록하는 단계이다. 그리고 등록된 사용자 개인 정보와 Top-N 알고리즘을 사용해 기존의 유사한 사용자 모임으로 분류한다.

사용자 행동분석 단계에서는 인기 있는 영화들을 사용자에게 임의로 추천한다. 그리고 사용자가 등록한 영화 평가를 분석하여 사용자 등급 행렬(User-Rating Matrix)을 생성한다. 또한, 생성된 사용자 등급 행렬을 기반으로 사용자에게 더 적합한 영화들을 추천한다.

SNS 연결 단계에서는 사용자가 등록한 SNS 계정에 영화를 추천하여 다른 사용자에게 추천된 영화를 홍보한다. 이와 같은 영화 추천 시스템의 구성은 그림 1과 같다.

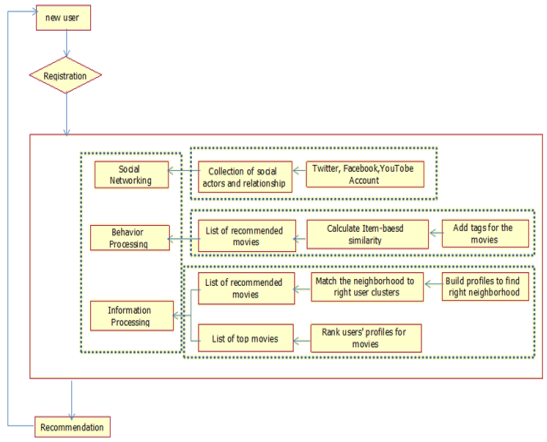

그림 1. 영화 추천 시스템의 구성도
Fig. 1 Configuration of Movies Recommendation System

영화 추천 시스템은 먼저 수집한 데이터를 분석하고 사용자 취향에 대한 데이터 모델을 구축한다. 그리고 사용자들의 유사도를 측정하고 순위를 계산한다. 영화 추천 시스템의 주요 처리 과정은 그림 2와 같다.

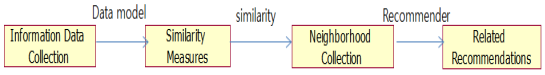

그림 2. 영화 추천 시스템 처리 과정
Fig. 2 Movie Recommendation System Processes

Ⅳ. 시스템 구현

4.1. 시스템 사용자 정보 수집

실험에 사용한 데이터는 Grouplens 사이트에서 제공하는 데이터를 사용하였다. 사용된 데이터는 2000년 6,040명의 MovieLens 사용자들이 3,900개의 영화에 1,000,209개의 평가한 데이터를 데이터베이스로 변환하여 사용하였다. 변환된 데이터베이스에서의 일부분인 사용자들의 성별과 연령, 직업, 관심 영화의 테이블은 그림 3과 같다.

시스템의 사용자 정보의 수집 화면은 그림 4와 같다. 영화 추천 시스템은 사용자가 제공한 정보를 분석하고 유사한 모임으로 분류한다. 그리고 사용자가 좋아할만한 영화를 추천한다.

Gender	Occupation			Age	Type	
Male	Administrator	Artist	Doctor	<18	Action	Adventure
Female	Educator	Engineer	Entertainment	18-25	Animation	Children's
	Executive	Healthcare	Homemaker	26-35	Comedy	Crime
	Lawyer	Librarian	Marketing	36-45	Documentary	Drama
	Programmer	Retired	Salesman	46-55	Fantasy	Film-Noir
	Scientist	Student	Technician	56-65	Horror	Musical
	Writer	Other		>65	Mystery	Romance
					Sci-Fi	Thriller
					War	Western

그림 3. 사용자 특정 분류
Fig. 3 User Certain Classification

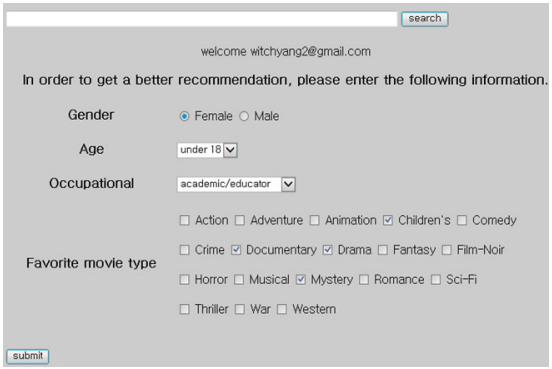

그림 4. 사용자 정보 수집
Fig. 4 User Information Collection

영화 추천 시스템에서는 사용자가 영화를 시청한 후 평가를 등록할 수 있다. 평가할 수 있는 영화는 15편이며 점수는 1부터 5까지 평가할 수 있거나 평가를 입력하지 않아도 된다. 구현 화면은 그림 5와 같다.

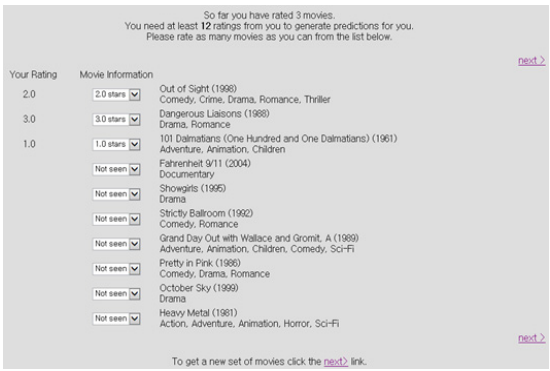

그림 5. 사용자 영화 평가 수집
Fig. 5 User Movies Evaluation Collection

4.2. Top-N 추천 알고리즘

새로운 사용자는 추천 시스템이 요구하는 정보들이 부족하여 Cold-Start 문제가 발생할 수 있다. 이러한 문제를 해결하는 방안으로는 Top-N 추천 알고리즘을 사용한다. 시스템은 데이터베이스에서 인기가 높은 영화의 순위를 정하여 새로운 사용자에게 영화를 추천한다. Top-N 추천 알고리즘의 가상코드는 그림 6과 같다.

```
#Input: users' preference list
#Output: list of recommended movies

1 for each movie in users' preference list for user
2         do put movie into list of recommended movies
3 end
```

그림 6. Top-N 추천 알고리즘
Fig. 6 Top-N Recommendation Algorithms

영화 추천 시스템에서 Top-N 추천 알고리즘을 사용하여 구현한 화면은 그림 7과 같다. 그리고 추천해준 영화가 사용자의 행동에 미치는 영향을 분석한 뒤 SNS을 사용하여 다른 사용자에게 연결한다.

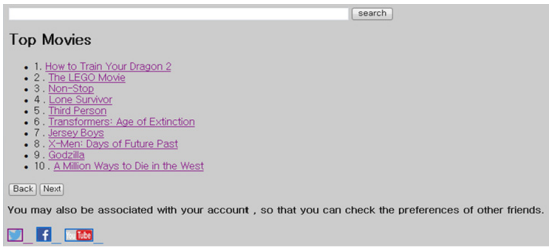

그림 7. Top-N 추천 알고리즘의 구현 화면
Fig. 7 Implementation Screen of Top-N Recommendation Algorithm

Ⅴ. 실 험

5.1. 유사 알고리즘

유사성을 계산할 때 자주 사용하는 알고리즘은 피어슨 상관관계(Pearson Correlation Similarity)와, 유클리드 거리(Euclidean Distance Similarity), 타니모토 계수(Tanimoto Coefficient Similarity), 로그우도(LogLikelihood Similarity)를 사용한다. 네 개의 알고리즘을 측정하여 그림 8과 같이 평균 절대 차이 값 결과를 측정하였다.

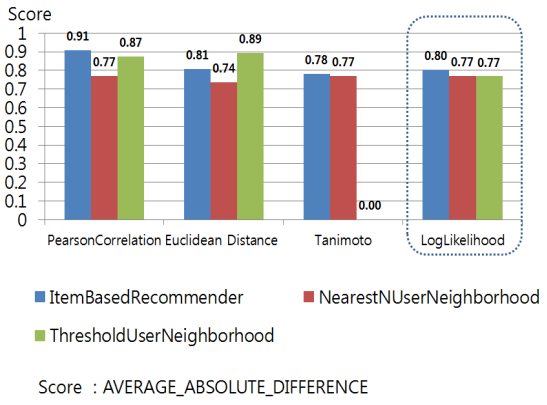

그림 8. 유사 알고리즘의 평균 절대 차이 결과
Fig. 8 Average Absolute Difference Result of Similarity Algorithms

5.2. 추천 알고리즘

추천 시스템에서 사용되는 추천 알고리즘은 트리 클러스터 기반 추천 알고리즘(TreeCluster)과 일반 항목 기반 추천 알고리즘(ItemCF), KNN 항목 기반 추천 알고리즘(ItemKNN), Slope-One 추천 알고리즘(Slope-One), 일반 사용자 기반 추천 알고리즘(UserCF), SVD 추천 알고리즘(SVD)들의 평균 절대 차이 값 결과를 측정한다. 측정한 결과는 그림 9와 같다.

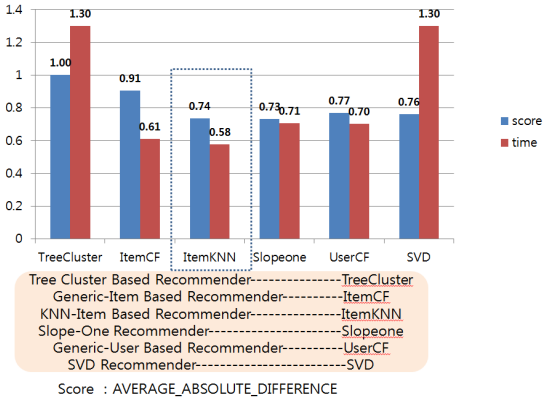

그림 9. 추천 알고리즘의 평균 절대 차이 결과
Fig. 9 Average Absolute Difference Result of Recommender Algorithms

5.3. 알고리즘 실험 평가

유사 알고리즘과 추천 알고리즘을 측정한 결과 로그우도 알고리즘과 KNN 항목 기반 알고리즘이 다른 알고

리즘에 비해 성능이 우수하다는 결과가 나왔다. 제안하는 영화 추천 시스템에서는 로그우도 알고리즘과 KNN 항목 기반 알고리즘을 사용하였으며 결과는 그림 10과 같다.

Hi, test3(test3@gmail.com), your id is 3

Recommended Movie

Movie Name	Published Year	Movie Type	Score
Cool Hand Luke	1967	Comedy Drama	5
Glory	1989	Action Drama War	5
Predator	1987	Action Sci-Fi Thriller	5
Hud	1963	Drama Western	5
King Kong	1933	Action Adventure Horror	5
Manchurian Candidate, The	1962	Film-Noir Thriller	5
Dirty Dozen, The	1967	Action War	5
Indian in the Cupboard, The	1995	Adventure Children's Fantasy	5
Aliens	1986	Action Sci-Fi Thriller War	5
Maverick	1994	Action Comedy Western	5
Full Metal Jacket	1987	Action Drama War	5
Queen Margot (La Reine Margot)	1994	Drama Romance	5
Get Shorty	1995	Action Comedy Drama	5
Heavy Metal	1981	Action Adventure Animation Horror Sci-Fi	5
Thin Blue Line, The	1988	Documentary	5
Heat	1995	Action Crime Thriller	5
Blade Runner	1982	Film-Noir Sci-Fi	5
Casablanca	1942	Drama Romance War	5
Hoop Dreams	1994	Documentary	5
Army of Darkness	1993	Action Adventure Comedy Horror Sci-Fi	5

Your Movies

Movie Name	Published Year	Movie Type	Score
Groundhog Day	1993	Comedy Romance	2
Rock, The	1996	Action Adventure Thriller	5
Ghost and the Darkness, The	1996	Action Adventure	4
Raising Arizona	1987	Comedy	4
Young Guns	1988	Action Comedy Western	5
Three Musketeers, The	1993	Action Adventure Comedy	4
Princess Bride, The	1987	Action Adventure Comedy Romance	5
Beverly Hills Ninja	1997	Action Comedy	3
Jurassic Park	1993	Action Adventure Sci-Fi	4
Bug's Life, A	1998	Animation Children's Comedy	5
Dragonheart	1996	Action Adventure Fantasy	4
Star Wars: Episode IV - A New Hope	1977	Action Adventure Fantasy Sci-Fi	5
Full Monty, The	1997	Comedy	2
Mask of Zorro, The	1998	Action Adventure Romance	4
Breakfast Club, The	1985	Comedy Drama	4
Star Wars: Episode V - The Empire Strikes Back	1980	Action Adventure Drama Sci-Fi War	4
Toy Story 2	1999	Animation Children's Comedy	3
Star Wars: Episode VI - Return of the Jedi	1983	Action Adventure Romance Sci-Fi War	4
Blade	1998	Action Adventure Horror	5
Butch Cassidy and the Sundance Kid	1969	Action Comedy Western	5
Stand by Me	1986	Adventure Comedy Drama	5
Mission: Impossible	1996	Action Adventure Mystery	3
Crocodile Dundee	1986	Adventure Comedy	4
Mummy, The	1999	Action Adventure Horror Thriller	2
28 Days	2000	Comedy	3
Raiders of the Lost Ark	1981	Action Adventure	5
Fish Called Wanda, A	1988	Comedy	5
Monty Python and the Holy Grail	1974	Comedy	5
Deliverance	1972	Adventure Thriller	4
Little Mermaid, The	1989	Animation Children's Comedy Musical Romance	4

그림 10. 추천 결과
Fig. 10 Recommender Result

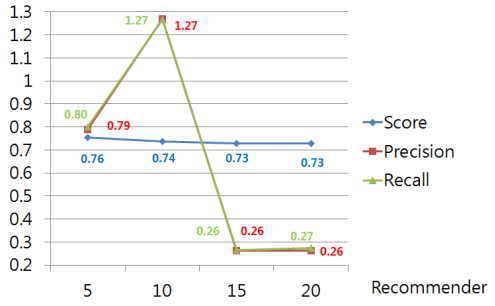

그림 11. 영화 추천 시스템의 정밀도와 재현율
Fig. 11 Precision And reproduction rate of Movies Recommendation System

추천 시스템의 성능은 보통 정밀도와 재현율로 평가할 수 있다. 그림 11은 KNN 항목 기반 알고리즘을 사용하여 제안하는 영화 추천 시스템의 성능을 측정한 결과다.

VI. 결 론

추천 시스템에서 사용되는 필터링 기법은 협업 필터링 기법과 내용 기반의 필터링 기법이 있다. 이러한 필터링 기법들은 장단점이 존재하고 잠재적인 문제점으로는 Cold-Start를 가지고 있다.

최근에 사용되는 추천 시스템에서는 Cold-Start 문제를 해결하는 방안으로 협업 필터링 기법과 내용 기반의 필터링 기법을 혼합한 필터링 기법을 사용한다. 혼합 필터링 기법은 장점을 부각하고 단점을 줄이는 필터링 기법으로 본 논문에서는 이러한 혼합 필터링 기법을 사용하는 영화 추천 시스템을 제안하였다. 제안하는 영화 추천 시스템에서는 유사도를 측정하는 유사 알고리즘과 추천에 사용되는 추천 알고리즘이 사용되며 우수한 알고리즘을 선별하기 위해서 실험을 진행하였다. 실험 결과 로그우도 알고리즘과 KNN 항목 기반 알고리즘이 다른 알고리즘에 비해 우수한 알고리즘으로 측정되었다. 제안하는 영화 추천 시스템은 로그우도 알고리즘과 KNN 항목 기반 알고리즘을 사용하여 구현하였으며 정밀도와 재현율을 통해 성능을 평가했다.

향후 연구로는 영화 추천 시스템의 성능을 향상시키기 위해 분산 처리 기법을 도입하는 연구가 필요하다.

REFERENCES

[1] Hornung, Thomas, et al, "Evaluating hybrid music recommender systems," *Web Intelligence (WI) and Intelligent Agent Technologies (IAT)*, vol. 1, 2013.

[2] Zhang, Zui, et al, "A hybrid fuzzy-based personalized recommender system for telecom products/services," *Information Sciences 235*, pp. 117-129, Jun. 2013.

[3] Sun, Mingxuan, et al, "Learning multiple-question decision trees for cold-start recommendation," *Proceedings of the sixth ACM international conference on Web search and data mining*, 2013.

[4] Ortega, Fernando, et al, "Improving collaborative filtering-based recommender systems results using Pareto dominance," *Information Sciences: an International Journal 239*, pp. 50-61, 2013.
[5] Van Meteren, Robin, and Maarten Van Someren, "Using content-based filtering for recommendation," *Proceedings of the Machine Learning in the New Information Age: MLnet/ ECML2000 Workshop*, 2000.
[6] Basilico, Justin, and Thomas Hofmann, "Unifying collaborative and content-based filtering," *Proceedings of the twenty-first international conference on Machine learning*, 2004.
[7] Sarwar, Badrul, et al, "Item-based collaborative filtering recommendation algorithms," *Proceedings of the 10th international conference on World Wide Web*, 2001.
[8] Wei, Suyun, et al, "Item-based collaborative filtering recommendation algorithm combining item category with interestingness measure," *Computer Science & Service System (CSSS)*, 2012.

정인용(In-Yong Jeong)

1997년 배재대학교 컴퓨터공학과(공학사)
2000년 배재대학교 컴퓨터공학과(공학석사)
2014년 ~ 배재대학교 컴퓨터공학과 박사과정
2011년 ~ 현재 (주)아이티어스 CTO
※ 관심분야 : 멀티미디어 문서정보처리, Big Data, 사물 인터넷

양새동(Yang, Xitong)

2013년 정주대학교 컴퓨터공학과(공학사)
2014년 ~ 배재대학교 컴퓨터공학과 석사과정
※ 관심분야 : Bigdata, Hadoop, Recommender System

정회경(Hoe-Kyung Jung)

1985년 광운대학교 컴퓨터공학과(공학사)
1987년 광운대학교 컴퓨터공학과(공학석사)
1993년 광운대학교 컴퓨터공학과(공학박사)
1994년 ~ 현재 배재대학교 컴퓨터공학과 교수
※관심분야 : 멀티미디어 문서정보처리, XML, SVG, Web Services, Semantic Web, MPEG-21, Ubiquitous Computing, USN, Bigdata