석사학위 논문

데이터 마이닝 기반 협업 필터링을 활용한 여행지 추천 기법

A Scheme for Recommending Locations to Tourists using Collaborative Filtering based on Data Mining

2017년 12월

숭실대학교 대학원

IT융합학과

이 태 린

석사학위 논문

데이터 마이닝 기반 협업 필터링을 활용한 여행지 추천 기법

A Scheme for Recommending Locations to Tourists using Collaborative Filtering based on Data Mining

2017년 12월

숭실대학교 대학원

IT융합학과

이 태 린

석사학위 논문

데이터 마이닝 기반 협업 필터링을 활용한 여행지 추천 기법

지도교수 정 윤 원

이 논문을 석사학위 논문으로 제출함

2017년 12월

숭실대학교 대학원

IT융합학과

이 태 린

이 태 린의 석사학위 논문을 인준함

심 사 위 원 장 유 명 식 인

심 사 위 원 노 동 건 인

심 사 위 원 정 윤 원 인

2017년 12월

숭실대학교 대학원

감사의 글

석사 학위가 끝날 때까지 여러 가지 방법으로 공헌 해주신 사랑하는 부모님에 대한 감사의 말을 전한 후, 내내 지도해 주신 정윤원 교수님께 감사의 말씀을 전하고 싶습니다. 교수님의 끊임없는 인도와 의견은 더욱 완성도 있는 논문을 쓸 수 있도록 큰 도움이 되었고, 조언과 안내는 저의 논문에 도움이 되었을 뿐만 아니라 향후 제 미래의 진로 방향을 확고히 하는데 지대한 영향을 미쳤습니다. 또한 재학 중 지원을 아끼지 않고 도와준 IT융합학과 덕분에 이 논문을 완료 할 수 있었습니다.

목 차

표 목 차

그 림 목 차

국문초록

데이터 마이닝 기반 협업 필터링을 활용한 여행지 추천 기법

이 태 린

IT융합학과

숭실대학교 대학원

최근 관광 관련 데이터가 증가하고 있으며 기업에서는 고객 선호도에 맞는 맞춤형 서비스를 활용한 효과적인 마케팅을 위해 고객의 특성을 식별하기 위해 노력하고 있다. 이를 위해 고객의 여행지 별 선호도 데이터를 기반으로 한 추천 시스템의 개발 및 선호도를 잘 측정할 수 있는 협업 필터링 모형의 구축이 필요하다. 본 연구에서는 협업 필터링을 통한 여행지 추천 기법을 제안하며 방문자 수 및 재방문 건수 등이 협업 필터링의 성능에 미치는 영향을 분석하였다. 분석 결과 여행 장소의 방문자 수 나 재방문 건수가 단순 고려된 선호도보다 방문자 수와 재방문 건수가 복합적으로 고려된 가중치를 평점으로 사용하였을 때 협업 필터링의 성능이 좋아짐을 알 수 있었다.

ABSTRACT

A scheme for recommending locations to tourists using collaborative filtering based on data mining

LEE, TAE-RIN
Department of it convergence
Graduate School of Soongsil University

Recently, the amount of data related with tourism has been increased and companies are focusing on identifying the characteristics of customers for the purpose of efficient marketing to provide customized services. To this end, the development of recommendation system based on customer's preference data for each traveling location and the development of collaborative filtering model to measure the preference well are needed. In this paper, a scheme for recommending locations based on collaborative filtering is proposed and the effect of the number of visitors and the number of multiple visits on the performance of the proposed collaborative filtering was analyzed. Analysis result shows that the performance of the collaborative filtering is improved if the well-designed weight

based on both the number of visitors and the number of multiple visits is used as score, instead of simply considering the number of visitors and the number of multiple visits.

제 1 장 서 론

1.1 연구의 배경 및 목적

오늘날 데이터의 양이 폭발적으로 증가함에 따라 데이터를 효과적으로 가공하는 기술도 나날이 발전되고 있다. 온라인 쇼핑몰에서는 여러 가지 추천 기법을 사용하여 사용자에게 특화된 상품을 추천하여 사용자들이 만족스러운 쇼핑을 할 수 있도록 도와주며 기업의 이익도 최대한 이끌어 내고 있다. 이러한 추천 시스템은 최근 산업 전 분야로 그 범위를 넓혀 가고 있다. 추천 시스템에서는 음악 및 음식점 추천, 방송 편성, 버스 노선, 마케팅 등 사용자 개인의 선호도에 맞는 정보를 제공하여 사용자의 관심을 유도한다.

본 논문의 대상인 여행 분야에는 아직 제대로 된 추천 시스템이 없고 주로 기업에 맞는 이익구조를 위해 유명한 관광지만을 추천하는 방식으로 사용자에게 추천이 되고 있다. 최근 개인 배낭여행, 가족여행이 크게 성장하고 있고, 주말 또는 연휴에 다양한 국내, 국외 여행이 이루어지고 있다. 현재 관광 플랫폼에서는 고객 개개인의 여행 지불 희망 금액, 여행 기간 고려하여 적절한 상품을 추천해주고 구매로 이어지는 방식으로 추천이 이루어지고 있는데 개인이 직접 자신의 선호도를 고려하여 자유롭게 여행지를 고를 수 있게 된다면 기업을 통하지 않고도 개인 선호도에 따른 여행지를 선택할 수 있게 될 것이다. 이러한 방법을 통해 사용자들로 하여금 과거 다녀왔던 여행지들에 대해 평가하고, 새로운 여행지를 보게 되면서 여행에 대한 욕구를 불러일으킬 수 있다. 또한, 개인에게 있어 그들에게 유사한 개인적 기호에 맞추거나 병행하여 다른 요소들이나

다른 개인적 분야에 기초한 결과를 활용하여 직접적 기호와 유사한 여행 장소를 결정하는 데에 적용될 수 있다.

본 연구에서는 관광 공공데이터를 기반으로 관광업에서 협업 필터링을 통해 추천 시스템을 개발하고자 할 때 행렬 데이터의 원소가 협업 필터링의 성능에 미치는 영향을 연구하고자 한다.

본 논문 2장에서는 협업필터링을 이용한 추천 방법을 여행 산업 안에서 활용 방안을 설명한다. 본 논문 3장에서는 협업필터링을 이용한 추천 알고리즘에 대해서 설명한다. 본 논문 4장에서는 구현된 시스템에 대한 실험 및 고찰을 논하며 제 5장에서 결론을 맺는다.

제 2 장 이론적 배경

2.1 추천 시스템과 여행

본 연구자가 파악하기로는 아직까지 여행 산업에서 추천 시스템이 제대로 적용된 경우는 별로 없다. 지금까지 여행 산업의 주된 주체는 항공권을 판매 시 부수적인 옵션을 제공하여 좀 더 쾌적한 여행을 할 수 있도록 서비스를 제공하는 기업이었기 때문에, 선호 여행지, 이미 대중화된 유명 나라의 여행지를 중심으로 여행 서비스가 제공 되었으며 이는 매년 매시기 비슷한 장소이다. 그러나 최근 여행자들은 인터넷과 텔레비전 등 대중매체를 통해 좀 더 다양한 여행지를 보게 되고 좀 더 개인화된 여행 장소 및 특별한 여행 장소를 가고 싶어 한다. 이를 위해 대중이 접할 수 있는 여행지를 개인 선호도에 맞춰 추천해줄 수 있다면 좀 더 다양한 선택이 가능해질 것이다. 직접적인 예로 2011년 이후로 급감되었던 일본을 방문하는 여행객의 수가 최근 들어 급증하고 있다. 국가 여행정보에 따르면 2017년 한국인 최고 순위 해외 여행지도 일본이었으며, 2018년 희망 해외 여행지 또한 일본이다. 이에는 일본의 다양한 여행 산업 발전을 위한 노력을 볼 수 있는데, 관광객에게 더욱 다양한 선택과 경험을 제공하며 합리적인 여행금액을 제시하였다.

반면 국내 여행을 하는 내국인과 외국인 방문자의 수는 점차 줄어가고 있고 한류를 제외한 한국만의 콘텐츠 또는 매력을 찾기 힘들기 때문이다. 한국인이 접하는 국내 여행지에 대한 정보는 나날이 많아지고 있는 반면 이러한 여행 정보를 외국인 유입을 위한 방법으로 제대로 쓰여 지지 않고 있다고 볼 수 있다. 늘어나는 데이터들을 기반으로 내국인, 외국

인 개개인의 성격에 맞는 여행지를 추천할 경우 접하는 긍정적인 요소가 많거나 개인의 선호도에 맞다면 여행으로 이어질 것이며 한국 관광 산업을 발전을 기대해 볼 수 있다.

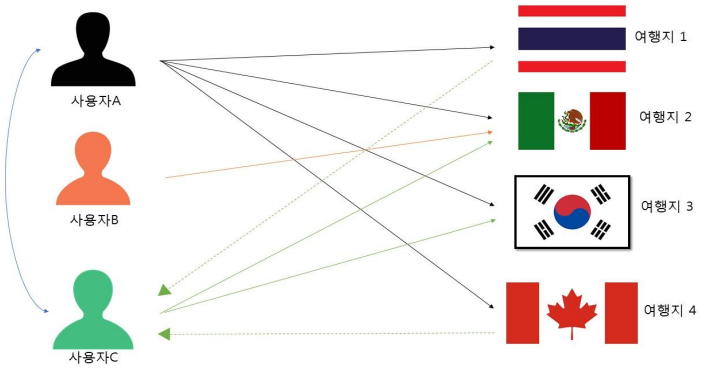
상기의 개별화된 여행지 추천을 위해서는 추천 시스템의 종류를 살펴볼 필요가 있는데 추천시스템의 종류는 크게 내용기반, 협업필터링, 하이브리드 추천시스템으로 분류할 수 있다. 내용기반과 같이 기존 여행지 정보만을 가지고 추천을 하는 경우 여행지에 관한 직관적인 정보만으로 추천 리스트가 만들어 졌다고 볼 수 있다. 그렇기에 사용자 개개인에 맞는 여행지 추천이 아닌 다수에게 대중적인 여행지 추천하는 방식이 되는 것이다. 협업필터링의 기본 개념은 비슷한 대상을 찾기 위해 비슷한 속성을 가진 데이터들을 비교, 분석 하는 것이다. 비슷한 속성을 가지거나, 비슷한 항목들을 비교하면 새 항목에 대해서도 유추가 가능할 것이라는 의도인 것이다. 이 방법을 여행지 추천에 도입을 하게 된다면 비슷한 성향을 가진 여행객들 끼리 그룹화 하여 비교, 분석 한 후 기존 여행지에 대한 만족도, 개인 성향 등을 기준으로 비슷한 성향이나 속성들을 비교하면 다음 여행지에 대한 추천이 가능한 것이다.

대표적인 협업 필터링에는 크게 사용자 기반과 아이템 기반이 있으며 각 방법의 장단점을 보완하여 결합한 방식을 하이브리드(Hybrid) 방식이라 한다. 이에 대한 상세한 설명은 아래와 같다.

2.1.1 사용자 기반 협업필터링(User-based Filtering)

사용자 기반 협력 필터링[1]은 사용자와 유사한 사용자를 추적하여 유사한 사용자가 선택한 아이템을 추천하는 방식이다. [그림 2-1]은 사용자기반 협업 필터링의 기본 개념도로 예를 들어, 사용자A가 여행지1, 여

행지2, 여행지3, 여행지4를 여행하였고, 사용자B는 여행지2, 사용자C는 여행지2와 여행지3을 다녀왔다. 사용자C에게 여행지를 추천하기 위해 같은 여행지를 다녀온 대상 사용자가 선정되면, 사용자 이력을 기반으로 사용자C와 같은 여행지를 다녀온 사용자와 유사도를 계산한다. 사용자B보다는 사용자A가 같은 여행지를 더 많이 다녀왔기 때문에, 사용자A가 다녀온 여행지1과 여행지4를 추천받게 된다. 사용 또는 구매한 아이템이 일치할수록 유사도가 높게 나타난다.



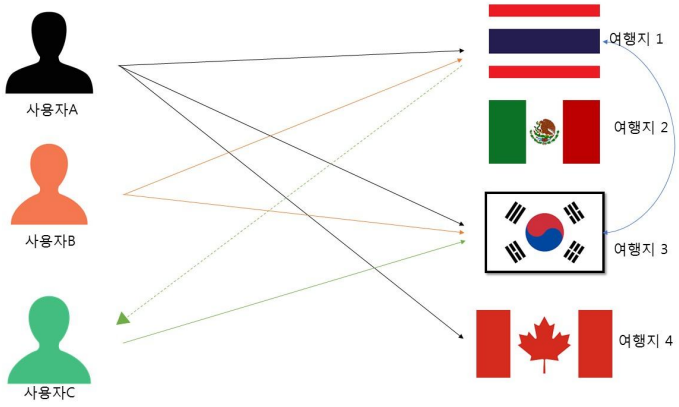
[그림 2-1] 사용자기반 협업 필터링 기본 개념도

2.1.2 아이템 기반 협업필터링(Item-based Filtering)

아이템 기반 협력 필터링[2]은 특정 여행지를 다녀온 사용자에게 다른 여행지를 추천하기 위하여 여행지끼리의 유사도를 비교한다. 그렇다면 대상 여행지를 다녀온 사용자 집단과 유사한 형태를 보이는 여행지가 존재할 가능성이 있으며, 이러한 정보를 기반으로 특정 사용자에게 여행지를 추천하는 것이 아이템 기반 협업 필터링의 기본 개념이다. [그림

2-2]는 아이템 기반 협업 필터링의 기본 개념도로 예를 들어 여행지3을 다녀온 여행자C에게 다른 여행지를 추천하기 위해서는 여행지3을 다녀온 다른 사용자를 찾는다. 그 뒤 모든 사용자들이 다녀온 각 여행지별 사용자 목록을 기준으로 유사도를 계산한다. 여행지3을 다녀온 여행자는 사용자A, B이다. 사용자A가 다녀온 여행지는 여행지1, 2, 3, 4이다 사용자B가 다녀온 여행지는 여행지1, 3이다. 이 경우 여행지3을 다녀온 모든 사용자들이 방문한 여행지 별 사용자 목록은 아래와 같으며 여행지3을 다녀온 여행자와 공통 여행자가 많은 여행지1이 추천된다.

여행지1 = {사용자A, 사용자B}
여행지3 = {사용자A, 사용자B, 사용자C}
여행지4 = {사용자A}



[그림 2-2] 아이템기반 협업 필터링 기본 개념도

2.1.3 Hybrid 협업 필터링(Hybrid-based Filtering)

하이브리드 협업 필터링은 사용자 기반 협업 필터링과 아이템 기반 협업 필터링의 단점을 보완 하여 제안하는 기법[3]으로, 사용자 기반 협력 필터링 기법에서 인구통계학적 데이터에 아이템 기반 협업 필터링 기법을 사용 할 때 얻는 아이템에 관련된 속성을 추가 하던가, 아이템 기반 협업 필터링 기법을 사용 중 얻을 수 있는 컨텐츠 정보에 사용자 프로필에서 찾을 수 있는 항목을 접목시켜 좀 더 능동적인 알고리즘을 만들어 낸다.

2.2 유사도 계산

협업 필터링에서 유사도를 계산하는 방법으로 사용자 기반 협업 필터링에서 많이 사용하는 피어슨 상관계수와 아이템베이스에서 많이 사용되는 코사인 유사도를 볼 수 있다.

2.2.1 피어슨 상관계수(Pearson's Correlation)

피어슨 상관계수(Pearson correlation coefficient)[4]는 두 변수간의 관련성을 구하기 위해 보편적으로 이용된다. 개념은 다음과 같다.

r = X와 Y가 함께 변하는 정도 / X와 Y가 따로 변하는 정도

r 값은 X 와 Y 가 완전히 동일하면 +1, 전혀 다르면 0, 반대방향으로 완전히 동일하면 -1 을 가진다. 결정계수 (coefficient of determination)

는 r^2 로 계산하며 이것은 X 로부터 Y 를 예측할 수 있는 정도를 의미한다.

$$Pearson = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \overline{y})^2}} \quad (2.1)$$

2.2.2 코사인 유사도(Cosine Similarity)

코사인 유사도[5]란 각 데이터의 두 벡터 간 각도의 코사인 값을 이용하여 측정된 벡터간의 유사한 정도를 의미한다. 각도가 0°일 때의 코사인 값은 1이며, 다른 모든 각도의 코사인 값 은 1보다 작다. 따라서 이 값은 벡터의 크기가 아닌 방향의 유사도를 판단하는 목적으로 사용되며, 두 벡터의 방향이 완전히 같을 경우 1, 90°의 각을 이룰 경우 0, 180°로 완전히 반대 방향인 경우 -1의 값을 갖는다. 1은 유사하다, 0은 관련이 없다, -1은 반대성향이다 라고 볼 수 있다.

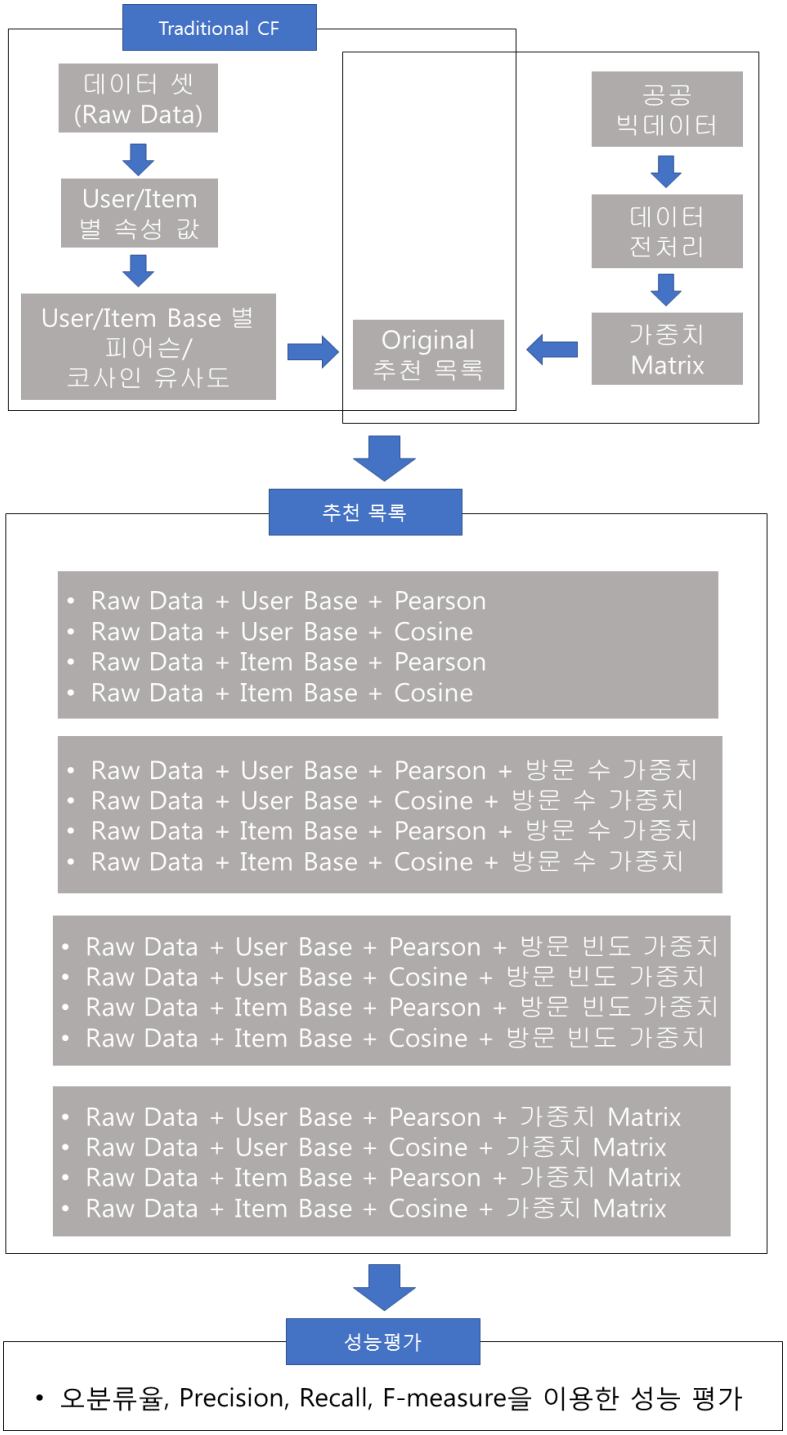
$$Cosine = \cos(\theta) = \frac{A \cdot B}{||A||||B||} = \frac{\sum_{i=1}^{n} A_i \times B_i}{\sqrt{\sum_{i=1}^{n}(A_i)^2} \times \sqrt{\sum_{i=1}^{n}(B_i)^2}} \quad (2.2)$$

제 3 장 제안 기법

3.1 시스템 설계

기존에 연구 되었던 협업 필터링 추천 시스템에서는 대부분 평가점수만을 기준으로 결과 데이터를 예측하였다. 그러나, 본 연구에서 제안하는 추천 시스템은 기존에 연구에서 활용해왔던 아이템에 대한 평가 점수 뿐만 아니라, 여행지별 방문 수, 방문 빈도수를 통계 및 수치화하여 도출해낸 가중치를 함께 적용하여 여행자 개인 선호도에 맞는 여행지를 추천하게 된다.

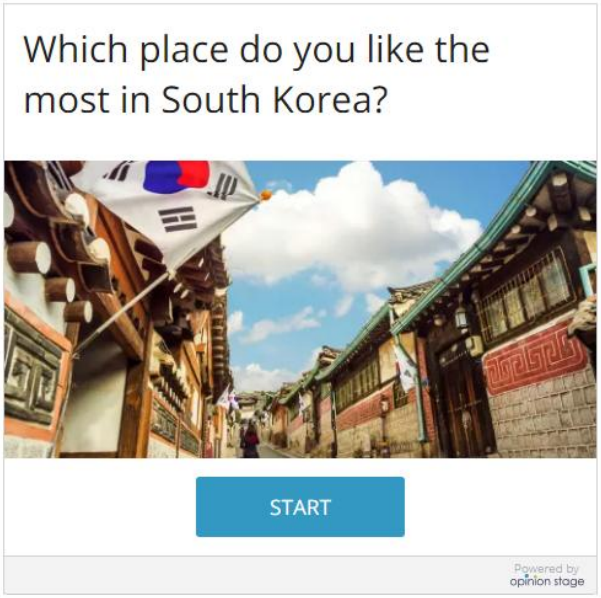
여행 선호도에 따른 방문 수, 방문 빈도수 데이터를 기반으로 공공 빅데이터 중 쇼핑을 선호 하는 사람들에게 추천 장소를 제안하기 위해서는 데이터 셋을 필요로 하게 되는데, 이는 설문조사를 통해 개인에게 특정한 장소를 보여주었을 때 반환되는 선호 점수 내역을 가진 Raw Data를 이용 할 수 있다. 이 데이터를 이용하여 User/ Item Base별 피어슨/코사인 유사도를 이용하여 비슷한 사용자를 찾아 낼 수가 있다. 또한 이 결과로 각 장소에 맞는 속성 값들을 분석하면 사용자가 어떠한 장소를 좋아하는지 알 수 있고, 그 속성에 맞는 빅 데이터를 전 처리하여 가중치를 부여한다. Row Data, 방문 수 가중치, 방문빈도 가중치, 가중치 Matrix 별로 오분류율, Precision, Recall, F-measure을 이용하여 성능을 평가하여 가장 효율적인 추천 방법을 도출해본다.



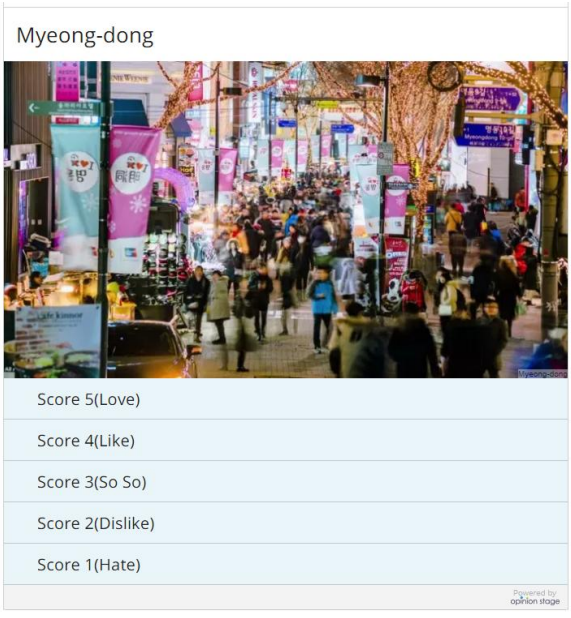
[그림 3-1] 여행지 추천을 위한 제안 기법 흐름도

3.1.1 Data set

본 연구에서는 협업 필터링에서 개인선호도에 맞는 여행지 추천을 위하여 각 장소와 사용자별 속성에 따른 가중치를 부여하는 기법을 제안하고자 한다. 이 실험을 위한 Dataset으로 한국 방문 시 각 장소별 선호도 온라인 조사로 얻은 254명의 데이터가 있다. 이를 이용하여 80%는 훈련용 데이터로, 20%는 테스트용 데이터로 이용하였다.



[그림 3-2] Survey 시작



[그림 3-3] Survey - Myeong Dong

[표 3-1] 협업필터링에서 사용하는 데이터 셋 예시

사용자	여행지	평가
1	10	1
1	1	2
1	2	5
1	3	5
1	6	5
1	7	1
1	8	5
2	5	5
2	6	4.5
2	7	1
2	8	5
3	1	2.5
3	2	4.5
3	3	4
3	4	3
3	7	4
3	8	5
4	10	5
4	1	5
4	2	5
4	6	1
4	7	4
4	8	1
5	2	4.5
5	3	4

3.1.2 가중치 부여 방법

개인 선호도에 맞는 가중치를 주기 위해서는 각 여행지 별 속성을 이용하여 개인선호도를 파악해야한다. 각 여행지 별 속성으로 쇼핑, 역사, 맛 집, 볼거리, 체험, 학습으로 특성을 나눌 수가 있는데, 이는 아래 [표 3-2]과 같다.

[표 3-2] 여행지 별 속성

	쇼핑	역사	맛 집	볼거리	체험	학습
여행지1	1	1	1	0	0	0
여행지2	1	1	0	1	1	0
여행지3	1	0	1	0	0	0
여행지4	1	1	1	0	0	1
여행지5	0	0	1	1	1	0
여행지6	1	1	0	1	1	1
여행지7	1	0	0	0	0	1
여행지8	0	0	1	0	1	0
여행지9	0	0	0	1	1	1
여행지10	1	1	0	1	0	0

[표3-2]을 기준으로 특정 사용자가 다녀온 여행지 목록을 보고 사용자의 개인 속성을 알 수 있다.

아래는 위의 [표 3-1] 데이터 셋을 각 사용자들이 다녀온 여행지로 표현한 행렬 표이다.

[표 3-3] 사용자 별 여행지 행렬표

	여행지1	여행지2	여행지3	여행지4	여행지5	여행지6	여행지7	여행지8	여행지9	여행지10
사용자1	1	1	1	0	0	1	1	1	0	1
사용자2	0	0	0	0	1	1	1	1	0	0
사용자3	1	1	1	1	0	0	1	1	0	0
사용자4	1	1	0	0	0	1	1	1	0	1
사용자5	0	1	1	0	0	0	0	0	0	0

예를 들어 사용자5는 여행지2, 3을 방문 하였는데, 여행지2, 3의 속성은 아래와 같다.

여행지2 = {1, 1, 0, 1, 1, 0}

여행지3 = {1, 0, 1, 0, 0, 0}

모든 여행지의 각 속성의 수치 합의 평균은 사용자의 속성과 같다고 할 수 있다.

$$\text{사용자의 쇼핑 선호 수치} = \frac{\text{여행지1의 쇼핑 수치} + \text{여행지2의 쇼핑 수치} + \text{여행지3의 쇼핑 수치}}{\text{사용자가 다녀온 전체 여행지 수}}$$

(3.1)

각 사용자가 다녀온 여행지별 각 속성의 합을 구하면 아래 표와 같이 사용자들의 속성을 나타 낼 수 있다.

[표 3-4] 사용자 별 속성

	쇼핑	역사	맛 집	볼거리	체험	학습
사용자1	0.8571	0.5714	0.4286	0.4286	0.4286	0.2857
사용자2	0.5000	0.2500	0.5000	0.5000	0.7500	0.5000
사용자3	0.8333	0.5	0.6667	0.1667	0.3333	0.3333
사용자4	0.8333	0.6667	0.3333	0.5	0.5	0.3333
사용자5	1	0.5	0.5	0.5	0.5	0

각 사용자의 속성은 순위가 생기는데 사용자5의 경우 쇼핑을 가장 선호한다고 볼 수 있다. 이런 경우 쇼핑을 위한 여행지에는 가중치를 줌으로써 개인성향에 따른 여행지 추천이 가능하게 된다.

속성별 여행지 가중치에 대한 수치는 3.3 데이터 전처리에서 설명한다. 3.3 데이터 전처리를 통해 얻은 방문 수, 방문빈도 수, 장소 별 방문 빈도수 가중치 Matrix까지 총 3가지 가중치 행렬표가 있다. 이 가중치를 각 아이템기반 필터링, 유저기반 필터링 별 피어슨 상관계수와 코사인 유사도 값 4개의 결과에 부여 해보고자 한다.

3.2 데이터 전처리

관광업에서 장소 별 가중치 부여 방법으로는 고객의 개인 데이터를 통해 가중치를 부여할 수 있는 방법은 때에 따라 다양하게 개발할 수 있지만, 본 연구에서는 여행지별 방문 데이터를 살펴보기로 한다. 여행자의 개인 데이터라고 할 수 있는 항목 중 쇼핑을 선호하는 여행자들을 위한 여행 장소 추천을 위하여, 쇼핑을 목적으로 여행지 장소 방문 데이터를 살펴본다.

쇼핑을 선호, 불호에 따라 특정 장소에 방문을 희망하는지 아닌지, 그 장소에 방문하였을 때 쇼핑 하고자 하는 욕구가 충족 되었는지, 다시 재방문 의사가 있는지가 여행 만족도에 영향을 미칠 것이다. 그 외에 선호하는 날씨, 선호하는 음식 등에 따라 여행 만족도에 영향을 미칠 수 있지만 이를 배제한 결과를 도출해 본다.

3.2.1 공공 데이터

본 연구에서는 서울시 공공데이터[6]에서 제공하는 서울시 23개구에 방문한 관광객들 12,003명을 대상으로 한 문화관광 공공데이터이다. 이 데이터를 통해 협업적 필터링 구축을 위한 가중치 행렬데이터의 원소로 사용할 최적의 변수를 탐색하였다.

협업필터링으로 각 여행자 별 선호도에 맞는 여행지 추천을 위해서는 어떠한 데이터를 사용 할 것인지 데이터 전처리 과정이 필요하다. 여행지 추천에 관련된 선호 사항으로 날씨, 음식, 목적, 나이, 국적, 동행, 기간, 지출 범위 등 기준을 잡을 수가 있다. 보다 결과를 명확하게 알기위해 날씨나 음식같이 한 장소에서도 얼마 던지 변할 수 있는 환경적 요소, 인적 요소 보다는 목적에 따른 여행지 추천을 고려하였다. 좀 더 구체적으로 각 여행자의 취향에 따라 목적을 세분화 시킬 수가 있다. 쇼핑, 고궁여행, 식도락 여행, 도보여행, 자동차 여행, 카페 투어 등 생각해 볼 수가 있는데, 서울 여행지 중 쇼핑으로 유명한 장소 들이 몇 군데 있으므로, 명확하게 쇼핑하기 좋은 곳인지 아닌 장소인지 객관적으로 판단 가능 한 장소들이 결과로 나올 수가 있다.

서울시 공공 데이터 중 유명 쇼핑 장소 기준으로 성별, 연령, 직업, 목적, 여행 형태, 방한 횟수 데이터를 제공 받을 수 있었다. 이 Raw 데이터를 바로 이용할 수도 있겠지만, 데이터 품질을 개선하거나, 데이터 분석에 적합한 형태로 변환 시킬 경우 분석에 소요되는 시간, 비용 등을 줄일 수 있다.

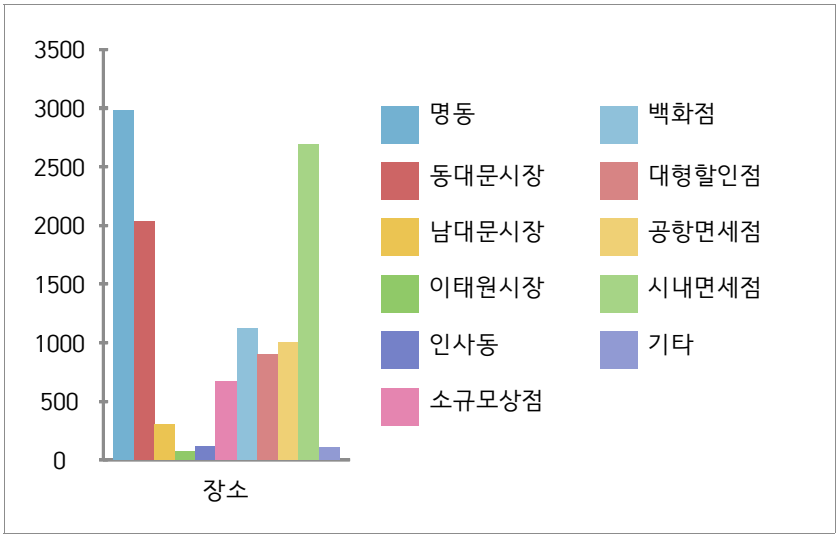
특정 사용자가 방문한 여행지의 종류가 매우 적거나 특정 여행지를 방문한 사용자 수가 매우 적으면 추천 시스템의 성능이 저하 될 수 있으므로[7], 여러 여행지 중 방문자 수가 최소 50건 이상의 장소를 선정 후 활용하여 추천 시스템을 구축하고, 이를 실험 및 평가하였다.

3.2.2 Raw Data

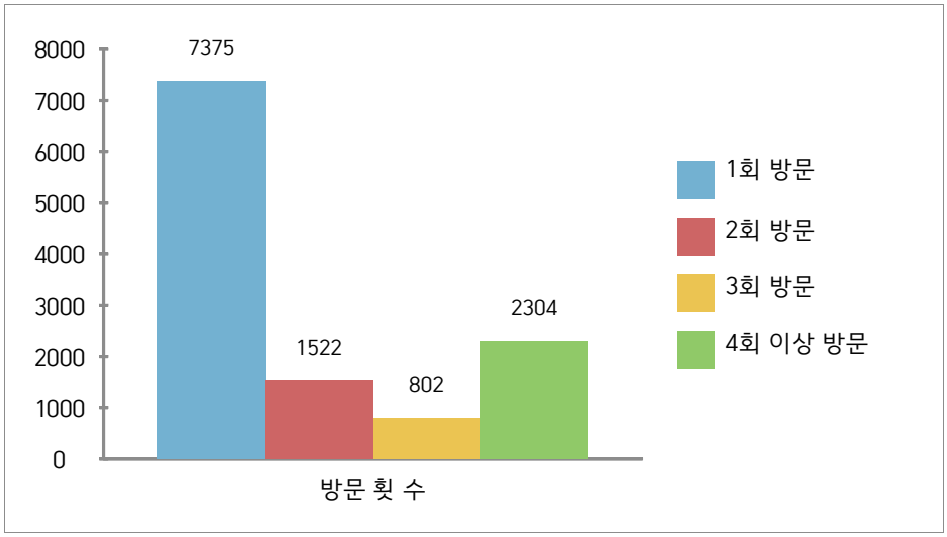
[그림 3-1] 에서는 쇼핑을 목적으로 할 경우 해당 장소를 방문한 여행자 수에 대한 그래프로 다양한 방문자 수의 데이터를 볼 수 있다. 각 장소에 방문한 여행자의 수는 각 지역 특성 별로 수치가 달라진다. 가장 높은 여행자 수가 있는 곳은 명동인데, 명동에는 면세점, 백화점, 아울렛, 프랜차이즈, 시장 등 대부분의 상권이 모여 있기 때문이다.

이 장소 외에도 어느 장소 든 물건을 판매하고 있기 때문에 계속해서 사용자 데이터를 수집 할 경우 협업 필터링의 가중치를 부여하게 되면, 가중치의 범위가 매우 넓어질 수 있다는 특징을 갖게 되고, 이론적으로는 0부터 무한대까지 부여할 수 있게 된다. 재방문 건수도 방문자 수와 마찬가지로, 지역 특성에 따른 데이터로 볼 수 있다.

쇼핑을 위해 재방문을 불러오는 상권에 의해 여행자의 방문수가 늘어날 경우, 판매 시장이 넓어지기 때문에 중요한 요소라고 볼 수 있다. 따라서, 가중치의 범위가 매우 넓어진다는 특징을 갖게 되고, 이론적으로 0건부터 무한대까지 가중치로 부여할 수 있다.



[그림 3-4] 장소 별 쇼핑을 목적으로 방문한 여행자의 수



[그림 3-5] 장소 별 쇼핑을 목적으로 여행자의 재방문 건수

3.2.3 Data Conversion

앞서 확인한 바와 같이 협업 필터링의 가중치 데이터로 쇼핑을 목적으로 해당 장소를 방문한 여행자들의 수와 재방문 건수의 Raw 값을 그대로 활용할 수도 있지만, 이를 변환하여 활용하는 것도 가능하다. 이와 관련하여 다양한 변환 방법들이 시도될 수 있지만, 방문지역과 재방문 건수의 백분위 수를 활용하여 기존의 방문지역과 방문 건수 값을 일반적인 가중치 형태인 백분율 형태로 변환하고자 한다.

방문 여행자의 수 및 방문 건수 수준이 많을수록 높은 퍼센트를, 방문 여행자의 수 및 방문 건수 수준이 적을수록 낮은 퍼센트의 값을 가지도록 Raw 값을 변환하였으며, 이와 같은 방법으로 쇼핑을 목적으로 각 장소에 방문한 여행자의 방문 횟수와 재방문 건수 가중치 결과를 확인하면 [표 3-1]과 같다.

장소별 재방문 건수도 장소 별 여행자 마찬가지로 변환할 수 있으며, 쇼핑을 위해 각 장소에 방문한 여행자들의 방문 건수 백분위 수에 따른 구간 화 결과를 확인하면 아래와 같다. 다만, 방문 건수의 경우 집계 기간이 짧거나, 여행지 특성상 방문 건수가 많지 않은 여행지의 경우 방문 건수의 표본 공간이 적어 10개의 구간이 모두 다른 방문 건수 값을 가지지 않을 수도 있는 것을 유의해야한다.

[표 3-5] 장소 별 쇼핑을 목적으로 방문한 여행자의 수 별 가중치

	여행자 수(단위: 명)	가중치(단위: %)
이태원 시장	73	0.6081
인사동	117	0.9747
남대문 시장	301	2.5075
소규모 상점	675	5.6231
대형 할인점	903	7.5225
공항 면세점	1005	8.3722
백화점	1122	9.3469
동대문 시장	2030	16.9110
시내 면세점	2685	22.3675
명동	2981	24.8334
기타	112	0.9330

[표 3-6] 장소 별 쇼핑을 목적으로 여행자의 재방문 건수 별 가중치

	여행자 수(단위: 명)	가중치(단위: %)
1회 방문	7375	10
2회 방문	1522	20
3회 방문	802	30
4회 방문 이상	2304	50

3.2.4 Matrix

협업 필터링의 행렬 데이터에 활용할 원소, 즉 여행자의 방문 가중치라고 할 수 있는 데이터를 여행자의 방문 선호 정도를 잘 나타낼수록 더욱 정확한 추천 시스템을 만들 수 있다. 온라인, 오프라인을 막론하고 관광업에서 취급하는 상품은 매우 다양하며 여행지에 대한 방문 시기에 따라 여행지 별로 다른 방문 특성을 보이게 된다. 예를 들어, 가격 측면으로는 고가의 여행지와 저가의 여행지라는 특성이 있으며 방문 시기로 보았을

때는 계절성 여행지와 사계절 여행지 특성이 있을 수 있다. 또한, 방문 건수의 측면으로 보았을 때는 잦은 방문 빈도수와 낮은 방문 빈도수 등의 특성으로 구분이 가능할 것이다.

여행지의 특성을 고려하면서 고객이 특정 여행지에 대해 가지는 선호도를 보다 정교하게 측정할 방법으로는 방문자 수 및 방문 빈도 Matrix를 활용해 볼 수 있다. 다음은 쇼핑을 목적으로 할 경우 해당 장소에 방문한 방문자 수 및 방문 빈도 Matrix이다. 방문자 수 별 방문 빈도에 따라 데이터가 분포되도록 [표 3-7]과 같이 작성하였다.

[표 3 - 7] 장소 별 쇼핑을 목적으로 방문자 수 및 방문 빈도 Matrix

	1회 방문	2회 방문	3회 방문	4회 이상 방문
이태원 시장	111	17	8	23
인사동	172	36	16	43
남대문 시장	234	58	20	69
소규모 상점	768	154	92	223
대형 할인점	723	153	91	284
공항 면세점	830	146	79	238
백화점	765	165	87	257
동대문 시장	765	184	91	315
시내 면세점	1526	271	120	282
명동	1393	307	172	458
기타	88	33	28	112
계	7375	1522	802	2304

위와 같은 Matrix를 작성하기 위해서는 방문한 여행자 수와 방문 건수의 데이터 값 구간 화 작업, 즉 그룹화를 거쳐야 한다. 그룹화 방법[8][9][10]에도 여러 가지가 있지만, 방문한 여행자 수와 방문 건수와 같은 숫자 형 변수를 구간 화하는 방법으로 백분위 수를 활용하여 기존의 값을 구간 화하고 이를 Matrix의 행과 열로 구성하였다.

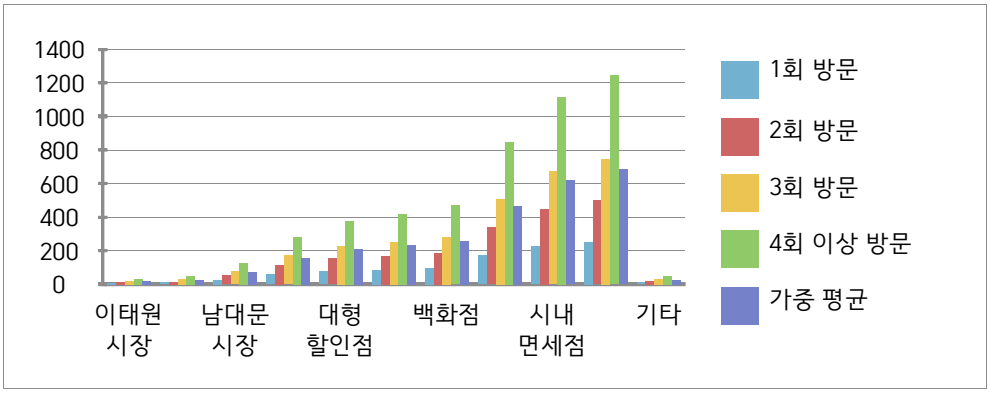
다음 단계로 Matrix를 구성하고 있는 데이터를 방문 수준에 따라 구분하고, 구분된 구역에 따라 가중치를 다르게 부여해야한다. Matrix의 오른쪽 아래로 향할수록 해당 여행지에 대한 방문 수준이 높다고 할 수 있으니 5점의 가중치를 부여하고, 왼쪽 위로 향할수록 방문 수준이 낮다고 할 수 있으니 1점의 가중치를 부여해 합리적인 가중치 부여 방법이라 생각해 볼 수 있다. 쇼핑을 목적으로 해당 장소를 방문한 여행자의 수와 재방문 건수 구간에 따라 Matrix에서 1점부터 5점까지의 가중치를 아래 표와 같이 부여해 보았다.

[표 3-8]과 같은 가중치 부여 방식에 따라, 2016년 1월부터 2016년 12월까지 쇼핑을 목적으로 명동을 방문한 여행자의 수가 많고 빈도수가 4회 이상 일 경우 가중치를 50을 부여받게 될 것이다. 반대로 쇼핑을 목적으로 할 시 방문자의 수가 적고 재방문 빈도수가 적은 이태원 시장의 경우 가중치 10을 부여받는다. 위 가중치 부여 방안에 따라, 쇼핑을 목적으로 방문 가능한 장소에 대해 부여 가능한 가중치 점수 별 여행자 분포는 아래와 같다.

위의 방법은 고객의 가치를 다양한 각도로 측정하는 고객가치평가 형태 중 하나인 고객 세분화의 방법과 같다고 할 수 있다. 표면적 데이터라 부를 수 있는 방문자 수를 기반으로 세분화 축을 선정하였고, 이 방법은 별도의 데이터 가공이 필요하지 않으므로 비교적 쉬운 세분화 방법의 하나로 활용되고 있다.

[표 3-8] 쇼핑을 목적으로 방문한 장소 별 가중치(단위: %)

	1회 방문	2회 방문	3회 방문	4회 이상 방문	가중 평균
이태원 시장	6.0813	12.1626	18.2439	30.4065	16.7236
인사동	9.7468	12.1626	29.2403	48.7338	24.9708
남대문 시장	25.0750	50.1500	75.2249	125.3749	68.9562
소규모 상점	56.2313	112.4625	168.6938	281.1563	154.6360
대형 할인점	75.2249	150.4499	225.6748	376.1246	206.8685
공항 면세점	83.7221	167.4442	251.1663	418.6105	230.2358
백화점	93.4688	186.9377	280.4065	467.3442	257.0393
동대문 시장	169.1103	338.2206	507.3309	845.5515	465.0533
시내 면세점	223.6754	447.3509	671.0263	1118.3772	615.1075
명동	248.3339	496.6678	745.0017	1241.6694	682.9182
기타	9.3302	18.6604	27.9907	46.6511	25.6581



[그림 3-6] 쇼핑을 목적으로 장소에 대한 가중치 별 방문자

제 4 장 실험 및 평가

본 연구에서는 관광업에서 확보 가능한 여행자 데이터를 기반으로, 협업적 필터링 구축을 위한 행렬 데이터의 원소로써 방문자 수, 재방문 건수, 방문자 수 별 구간 화, 재방문 건수 별 구간 화, 방문자 수 및 재방문 건수 기반 구간 화 수치를 사용, 각 가중치 부여 방법별로 추천 시스템을 구축하여 추천시스템 예측에 관한 실험해 보았으며 추천 시스템에 대한 성능을 평가하기 위해 지표를 만들어 확인해 보았다.

4.1 평가 척도

협업 필터링을 사용함으로써 발생 가능 한 문제들[11]이 몇 가지 있는데, 그 중 대표적으로 Cold-Start, 희소성이 있다. Cold-Start란 해당 유저의 평가 내역이 없는 경우 유사한 사용자를 찾을 수 없는 부분이고, 희소성이란 해당 사용자와 비슷한 사용자를 찾을 수 없는 경우이다.

이러한 문제들을 해결하여 추천 시스템의 성능을 평가하기 위해 12,003명의 여행자를 8:2의 비율로 구분하여 80%에 해당하는 9602명의 여행자를 훈련용 데이터, 20%에 해당하는 2401명의 여행자를 평가용 데이터로 구분하였다. 또한, 본 연구의 추천 시스템은 여행지 중 11개 여행지 장소 중 5개만 가지고 나머지 6개 여행지 중 방문이 예상되는 2가지 여행지를 추천하도록 작동하는 형태가 되도록 구축하였다.

본 연구의 실험은 방문 혹은 비방문과 같이 분류가 목적인 실험들이므로, 오분류율(Misclassification rate)을 기반으로 평가가 이루어지며, 추천 시스템의 성능을 평가하는 지표[12]로써 재현율(Recall), 정확률(Precision, Positive predictive value, PPV)[13], F-measure을 사용하였

다.

정확률(Precision)은 고객에게 추천된 여행지 중, 실제로 고객이 방문한 여행지의 수로 계산된다. 추천된 N개 여행지에 대한 정확성을 평가하기 위한 방법이다.

$$\mathrm{P}recision = \frac{tp}{tp + fp} \tag{3.1}$$

재현율(Recall, True positive rate, TPR)은 고객이 실제로 방문한 여행지 중 추천 시스템을 통해 실제로 방문한 여행지의 개수로 계산된다.

F-measure는 정확률과 재현율 두 개의 평가지표를 혼합하여 하나의 값으로 산출되는 지표이다.

$$Recall = \frac{tp}{tp + fn} \tag{3.2}$$

추천 여행지 수가 많아질수록 재현율은 증가하지만, 정확률은 감소하게 되므로, 좀 더 합리적인 평가를 위해 분류의 성능 및 효율성을 평가하는 지표로 F-measure가 사용될 수 있다.

$$F-measure = \frac{2 \times \mathrm{P}recision \times Recall}{\mathrm{P}recision + Recall} \tag{3,3}$$

F-measure는 0과 1 사이 값을 가지며, 0에 가까울수록 정확률과 재현율 중 하나가 상대적으로 낮다는 것을 의미하며, 반대로 1에 가까울수록 정확률과 재현율 모두 높다는 것을 의미한다.

또한 가장 일반적으로 분류 성능을 평가할 수 있는 척도로 오분류율을 사용할 수 있으며, 오분류율은 추천된 여행지와 추천되지 않은 여행지 모두를 옳게 분류한 비율을 의미한다.

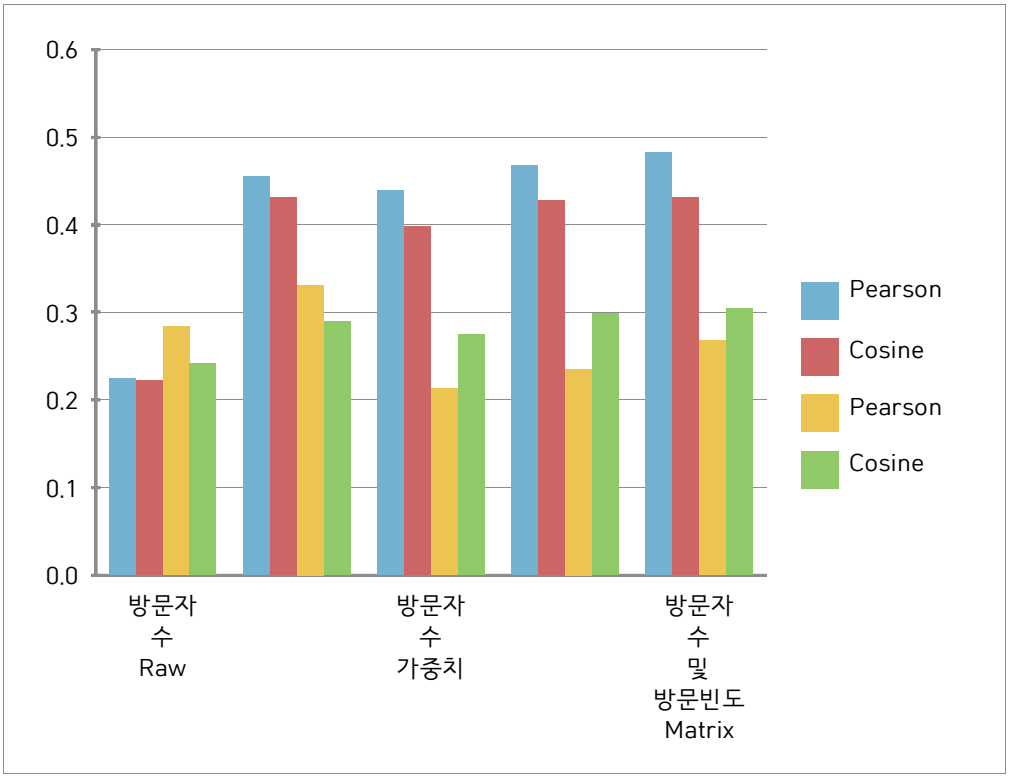
4.2 비교 결과

User-based filtering과 Item-based filtering 방법 별로 피어슨 상관계수 유사도와 코사인 유사도를 모두 고려하여 총 4가지 알고리즘인 User-based filtering의 피어슨 상관계수 유사도 측정 방법, User-based filtering의 코사인 유사도 측정 방법, Item-based filtering의 피어슨 상관계수 유사도 측정 방법, Item-based filtering의 코사인 유사도 측정 방법 별로 4가지 행렬표에 대한 최적의 결과를 갖는 조합 갖을 수 있도록 평가 지표들을 확인해 보았다.

다음은 4가지 행렬표에 대한 정확률 평가지표 비교이다.

[표 4-1] 행렬표 별 정확률 평가 결과표

구분	협업 필터링 방법	UBCF(사용자 기반 협업필터링)		IBCF(아이템 기반 협업필터링)	
	유사도 측정	Pearson	Cosine	Pearson	Cosine
가중치 부여 방법	방문자 수 Raw	0.2241	0.2219	0.2831	0.2410
	방문빈도 Raw	0.4549	0.4312	**0.3312**	**0.2889**
	방문자 수 가중치	0.4387	0.3976	0.2128	0.2745
	방문빈도 가중치	0.4672	0.4275	0.2355	0.2978
	방문자 수 및 방문빈도 Matrix	**0.4821**	**0.4311**	0.2678	0.3048



[그림 4-1] 행렬표 별 정확률 평가 결과 비교

사용자기반 협업 필터링의 피어슨 상관계수 유사도 측정방법에서는 방문자 수 및 방문빈도 Matrix가 0.4821로 정확도가 가장 높았으며, 사용자기반 협업 필터링의 코사인 유사도 측정방법에서도 방문자 수 및 방문빈도 Matrix가 0.4452으로 가장 높았다.

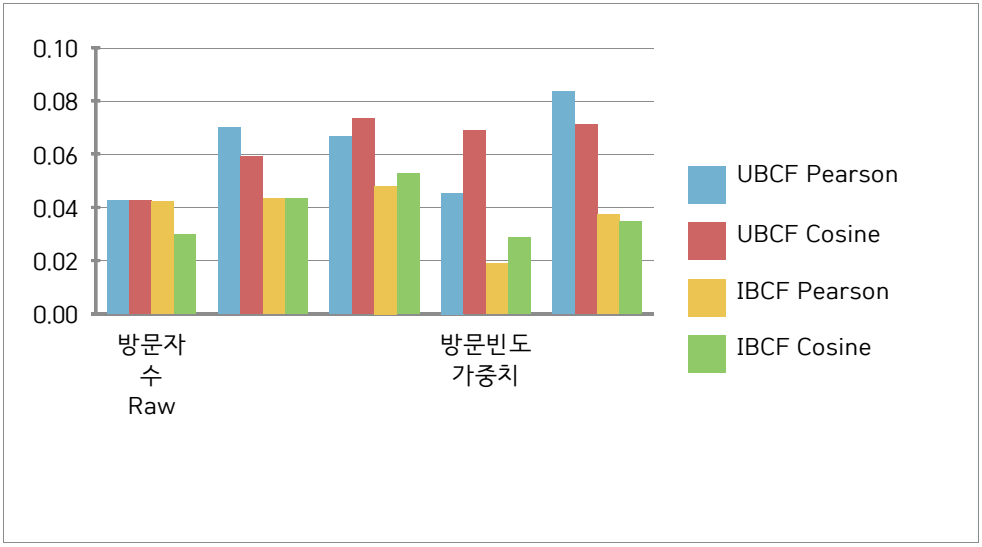
마찬가지로, 아이템기반 협업 필터링의 피어슨 상관계수 유사도 측정방법에서는 방문 건수 Raw에 의한 가중치 부여 방법이 0.3312으로 가장 높았고, 아이템기반 협업필터링의 코사인 유사도 측정 방법에서도 방문건수 Raw에 의한 가중치 부여방법의 정확률이 0.2889으로 가장 높았다.

협업 필터링 알고리즘 조합 방법에 따라 가중치 부여 방법별 정확률 순위가 다르지만, 정확률이 0.4821인 사용자기반 협업필터링과 피어슨 상관계수 유사도 측정방법 조합의 방문자 수 및 방문 빈도 Matrix 가중치 부여 방법이 가장 좋은 정확률을 나타내고 있다.

다음은 행렬표 별 재현율 평가지표 비교이다.

[표 4-2] 행렬표 별 재현율 평가 결과표

구분	협업 필터링 방법	UBCF(사용자 기반 협업필터링)		IBCF(아이템 기반 협업필터링)	
	유사도 측정	Pearson	Cosine	Pearson	Cosine
가중치 부여 방법	방문자 수 Raw	0.0428	0.0427	0.0421	0.0297
	방문빈도 Raw	0.0699	0.0591	0.0435	0.0433
	방문자 수 가중치	0.0667	0.0734	**0.0481**	**0.0525**
	방문빈도 가중치	0.0453	0.0687	0.0188	0.0287
	방문자 수 및 방문 빈도 Matrix	**0.0834**	**0.0711**	0.0371	0.0345



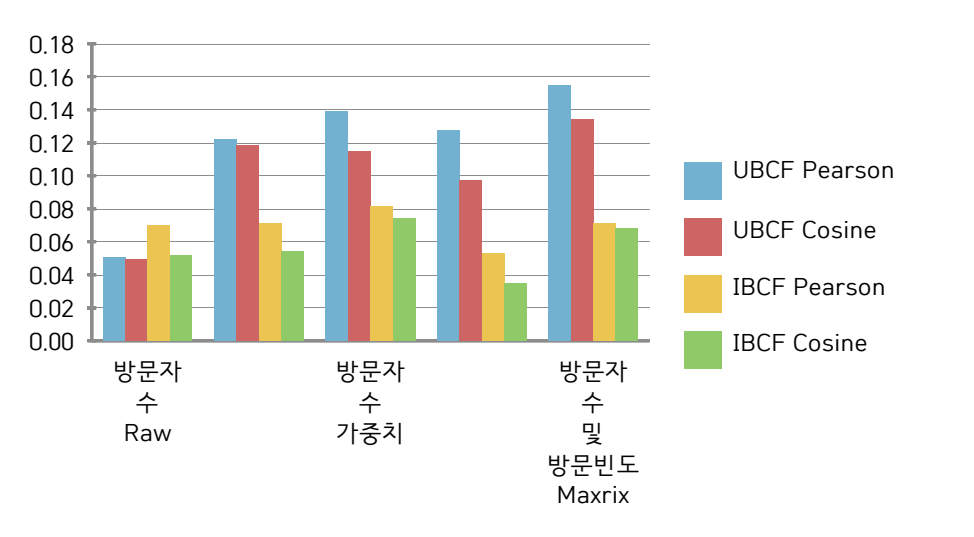
[그림 4-2] 행렬표 별 재현율 평가 결과 비교

협업 필터링의 2가지 방법과 유사도 측정의 2가지 방법을 조합한 총 4가지 방법에 따라, 5가지 가중치 부여 방법별 재현율을 비교하여 보았다. 정확률과 마찬가지로 협업 필터링 구축 4가지 방법 별 순위는 다르지만, 재현율이 가장 높은 것은 사용자기반 협업필터링의 코사인 유사도 측정 방법 조합과 방문자 수 및 방문 빈도 Matrix 가중치 부여 방법이 가장 좋은 재현율(0.0834)을 나타내고 있다.

[표 4-3]은 행렬표 별 F-measure 평가지표 비교이다. F-measure 평가지표 또한, 협업 필터링 구축 4가지 방법별 순위는 다르지만, F-measure 값이 가장 높은 것은 사용자기반 협업필터링의 코사인 유사도 측정 방법 조합과 방문자 수 및 방문 빈도 Matrix 가중치 부여 방법이 가장 좋은 F-measure 값(0.1546)을 나타내고 있다.

[표 4-3] 행렬표 별 F-measure 평가 결과표

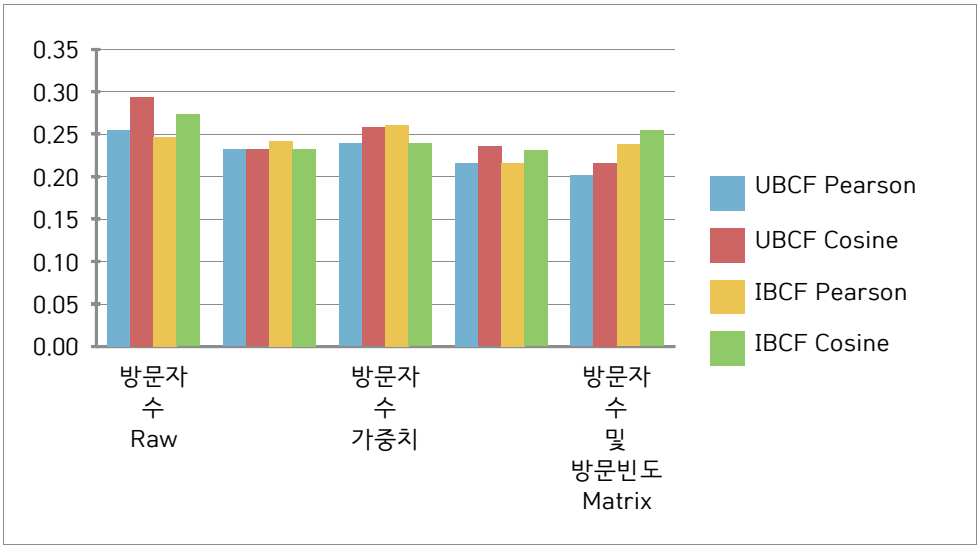
구분	협업 필터링 방법	UBCF(사용자 기반 협업필터링)		IBCF(아이템 기반 협업필터링)	
	유사도 측정	Pearson	Cosine	Pearson	Cosine
가중치 부여 방법	방문자 수 Raw	0.0503	0.0493	0.0703	0.0521
	방문빈도 Raw	0.1223	0.1181	0.0711	0.0544
	방문자 수 가중치	0.1387	0.1148	**0.0813**	**0.0745**
	방문빈도 가중치	0.1271	0.097	0.0532	0.0351
	방문자 수 및 방문빈도 Maxrix	**0.1546**	**0.1342**	0.0708	0.0678



[그림 4-3] 행렬표 별 F-measure 평가 결과 비교

[표 4-4] 행렬표 별 오분류율 평가 결과표

구분	협업 필터링 방법	UBCF(사용자 기반 협업필터링)		IBCF(아이템 기반 협업필터링)	
	유사도 측정	Pearson	Cosine	Pearson	Cosine
가중치 부여 방법	방문자 수 Raw	0.2549	0.2933	0.2464	0.2731
	방문빈도 Raw	0.2318	0.2317	0.2423	0.2319
	방문자 수 가중치	0.2388	0.2578	0.2607	0.2388
	방문빈도 가중치	0.2157	0.2355	**0.2157**	**0.2307**
	방문자 수 및 방문빈도 Matrix	**0.2011**	**0.2157**	0.2375	0.2540



[그림 4-4] 행렬표 별 오분류율 평가 결과 비교

오분류율은 정확률, 재현율, F-measure와는 반대로 낮을수록 추천 시스템의 성능이 좋음을 의미한다. [표4-4]는 행렬표 별 오분류율 평가지표 비교이다. 오분류율이 가장 낮은 것은 사용자기반 협업 필터링의 피어슨 유사도 측정 방법 조합과 방문자 수 및 방문 빈도 Matrix 가중치 부여 방법이 가장 낮은 오분류율(0.2011)을 나타내고 있다. 지금까지 살펴본 4가지 평가지표가 가장 우수한 협업 필터링 구축 방법 조합과 가중치 부여 방법을 정리해보면 아래와 같다.

[표 4-5] 평가지표 별 조합, 가중치 부여방법

평가 지표	평가 결과	협업 필터링 구축 조합		가중치 부여방법
		방법	유사도 측정	
정확률	0.4821	UBCF	Pearson	Matrix
재현율	0.0834	UBCF	Pearson	Matrix
F-meausre	0.1546	UBCF	Pearson	Matrix
오분류율	0.2011	UBCF	Pearson	Matrix

4가지 평가지표에 의해 가장 우수한 성능을 보인 협업 필터링 구축 조합은 때에 따라 사용자기반 협업 필터링의 피어슨 상관계수 유사도 측정 방법이 될 수 있다. 가중치 부여 방법은 4가지 평가지표 모두 방문자 수 및 방문 빈도 Matrix 가중치 부여 방법이 다른 가중치 부여 방법보다 상대적으로 우수한 방법으로 나타났다.

따라서, 관광업 데이터 기반 여행지 추천시스템을 개발하고자 할 때, 가중치를 부여하는 방법은 방문자 수 또는 재방문 건수를 그대로 사용하거나 백분위 수를 이용한 단순 구간 화 변환 방법보다, 방문자 수 및 재방문 건수에 기반을 둔 Matrix를 활용하여 가중치를 부여하는 것이 성능의 향상시킬 수 있다고 보여 진다.

제 5 장 결론

온라인과 오프라인을 막론하고 고객이 방문하기를 원하는 여행지를 최대한 빠르고 정확하게 파악, 예측하여 적시에 추천해주는 것이 고객 관계 관리 업무의 효율성을 증대시키고 방문 고객의 만족도를 향상할 수 있으며 이를 통해 개인 여행자들의 선택을 폭을 넓혀 다양한 선택을 불어 일으킬 수 있다.

본 연구에서는 관광업의 여행지 별 방문 데이터를 활용하여 여행지 추천 시스템을 구축할 때, 협업 필터링 행렬 데이터의 원소, 즉 가중치로 어떤 데이터를 활용해야 가장 좋은 성능의 추천 시스템을 구축할 수 있는지 비교 연구해 보았다. 서울시의 공공데이터를 이용하여 실험해 본 결과, 여행지 별 방문자 수나 재방문 건수, 방문자 수 별 구간 화 변환 방법, 재방문 건수 별 구간 화 변환 방법보다 방문자 수 및 재방문 건수가 모두 고려된 Matrix 방법이 가장 우수한 결과를 나타냈다.

이와 같은 결과가 나타나는 이유는 앞선 가중치 부여 방법들보다 해당 목적별 방문자 수와 재방문 건수 모두 고려되었기 때문에 방문 고객이 특정 지역에 대해 가지는 선호도를 보다 정교하게 측정할 수 있었기 때문이다.

정확한 방문 예측을 위해 관광업에서 여행자가 여행 사후에 매긴 해당 여행지에 대한 선호도 또는 만족도 평점은 큰 변수가 되기 때문에 더욱

다양한 평가지표를 만드는 것이 중요하다. 그러나 여행지 방문 사후의 선호도 정보는 일반적으로 수집하기 어려울 뿐만 아니라 수집하더라도 비용과 시간이 발생할 수 있다.

여행자가 의사 표현한 평점만큼 정확하고 정교한 가중치 부여를 위해, 본 연구에서 제안했던 것처럼 목적별 방문자 수, 재방문 건수의 2가지 변수 조합과 같이 여행지별 방문 데이터 내역에서 획득 가능한 정보를 복합적으로 고려해야 여행지 추천 시스템이 향상 될 것으로 보인다.

다만, 본 연구의 추천 단위는 최하위 동별 여행지의 추천이 아닌 목적별 대표 항목에 대한 추천이므로, 이 점을 고려하여 본 연구에서 제시한 Matrix에 의한 가중치 부여 방법을 활용해야 할 것이다. 따라서, 향후 과제로 각 목적 별 좀 더 구체적인 항목을 적용하여 보다 정교한 가중치 부여 방법이 연구되어야 할 것 이다.

참고문헌

[1] J. Hauke and T. Kossowski, “comparison of values of pearson′s and spearman′s correlation coefficients on the same sets of data,” Quaestiones Geographicae, 2011, vol. 30, issue 2, pages 87-93, 2011

[2] B. Sarwar, G. Karypis, J. Konstan, J. Riedl, "Item-Based Collaborative Filtering Recommendation Algorithms", Proc. 10th Int′l WWW Conf., 2001.

[3] Ya-Yueh Shih1 and Duen-Ren Liu, “Hybrid recommendation approaches: collaborative filtering via valuable content information,” Hawaii International Conference on System Sciences, 2005

[4] Benesty, Jacob, et al. "Pearson correlation coefficient." Noise reduction in speech processing. Springer Berlin Heidelberg, 2009. 1-4.

[5] Mihalcea, Rada, Courtney Corley, and Carlo Strapparava. "Corpus-based and knowledge-based measures of text semantic similarity." AAAI. Vol. 6. 2006.

[6] 공공데이터포털, “여행”, https://www.data.go.kr/search/index.do, (2017.08.15)

[7] 이홍주, 김종우, 박성주, “협업 필터링 기반 상품 추천에서의 평가 횟수와 성능”, 韓國經營科學會誌, 2006

[8]김정우, 박광현, "협업 필터링과 빈발 패턴을 이용한 개인화 그룹추천," The Journal of Korean Institute of Communications and Information Sciences '16-07 Vol.41 No.07, 2016

[9] 이오준, 유은순, “추천 시스템의 성능 안정성을 위한 예측적 군집화 기반 협업 필터링 기법”, J Intell Inform Syst 2015 March: 21(1):

119~142
[10] 이재웅, 이종욱, “상위 N개 항목의 추천 정확도 향상을 위한 효과적인 선호도 표현방법”, Journal of KIISE, Vol. 44, No. 6, pp. 621-627, 2017. 6
[11] 권형준, 홍광석, “잠재적 속성 선호도를 이용한 협업 필터링의 데이터 희소성 문제 개선 방법”, 인터넷정보학회논문지, 2013
[12] 조영빈, 조윤호, “구매순서를 고려한 개선된 협업필터링 방법론”, 한국지능정보시스템학회논문지, 2007
[13] Dr. Sergey Morozov, Dr. Xiaohui Zhong, “The Evaluation of Similarity Metrics in Collaborative Filtering Recommenders”, 2013 Hawaii University International Conferences, June 10th to June 12th