

머신러닝 기반 음식점 추천시스템 설계 및 구현

최환석¹ · 팡전² · 이우섭^{3*}

¹한밭대학교 멀티미디어공학과 박사후연구원, ²석사과정, ³한밭대학교 정보통신공학과 교수

Design and Implementation of the Machine Learning-based Restaurant Recommendation System

Hoan-Suk Choi¹ · Qian Peng² · Woo-Seop Rhee^{3*}

¹PostDoc, ²Master's Course, Department of Multimedia Engineering, Hanbat University, Daejeon 34158, Korea

³Professor, Department of Information Communication Engineering, Hanbat University, Daejeon 34158, Korea

[요 약]

최근 식습관 변화와 인터넷의 발전으로 앱 기반 음식주문 서비스가 생겨났고 이에 따라 배달 시장이 커지고 있다. 또한 사용자는 자신이 선호하는 대상을 쉽게 찾길 원하기 때문에 다양한 분야에 추천 시스템이 도입되고 있다. 추천시스템은 다양한 속성을 포함한 충분한 데이터가 확보되어야 안정적이며 신뢰적인 성능을 보인다. 따라서 본 논문에서는 다양한 목적 및 상황에 맞는 추천을 제공하기 위해 데이터 속성 확장 메커니즘과 머신러닝 기반 음식점 추천시스템을 제안한다. 제안하는 시스템은 추천 리스트 생성을 위해 자연어 평가 기반 음식점 특징 추출 기법, K-means 기반 평가점수 유사그룹 생성기법, 선형회귀 기반 주문량 예측 기법을 제공한다. 본 논문은 Python을 기반으로 제안하는 시스템을 구현하였으며 중국 광저우 지역의 음식점 평가 데이터를 기반으로 추천목록 생성 실험을 수행하였다. 제안하는 시스템은 동적인 특징 추출로 대상의 특성을 유연하게 반영할 수 있으며, 보다 상세한 사용자의 취향 구분이 가능하다. 또한 주문량 예측결과 20% 미만의 오차를 보여 추세파악이 가능할 것으로 생각된다.

[Abstract]

Recently, due to the changes in dietary life style and developments of the Internet technology, the app-based food ordering services have been emerged, and the delivery market continues to grow. Also, recommendation system is being introduced in a variety of areas because users want to easily find preferred targets. A recommendation system requires sufficient data including various attributes to provide stable and reliable performance. Thus, this paper proposes the data feature extension mechanism and machine learning based restaurant recommendation system to provide recommendations various purposes and situations. The proposed system provides the natural language comment based restaurant feature extraction method, the K-means based review score similar group generation method and the linear regression based ordering quantity prediction method to generate recommendation lists. Also, this paper implemented the proposed system by Python and conducted experiment of generating recommendation list based on restaurant review data of Guangzhou. The proposed system can flexibly reflect the feature of the target and allow for more detailed user taste. Also, it can predict the order trend because the order quantity prediction shows a margin of error is less than 20%.

색인어 : 추천시스템, 음식점 추천, 중국어 평가의견 데이터, 머신러닝, 케이-평균

Key word : Recommender system, Restaurant recommendation, Chinese comment data, Machine learning, K-means

<http://dx.doi.org/10.9728/dcs.2020.21.2.259>



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.

Received 02 December 2019; **Revised** 10 January 2020

Accepted 25 February 2020

***Corresponding Author; Woo-Seop Rhee**

Tel: +82-42-821-1749

E-mail: wsrhee@hanbat.ac.kr

I. 서론

통신기술의 발달로 언제 어디서나 인터넷에 접속할 수 있는 시대가 되면서 수많은 콘텐츠와 서비스가 제공되고 있다. 이러한 정보의 홍수 속에서 사용자는 자신이 원하는 정보를 더욱 쉽게 찾고자 한다. 따라서 다수의 서비스에 추천 시스템을 도입하고 있다. 대표적인 추천 시스템은 아마존의 '북 매치'이다. 이는 사용자 구매 이력 및 평점 데이터를 바탕으로 도서를 추천한다. 추천 시스템으로 인해 아마존은 폭발적인 성장을 이뤄냈고, 판매량의 35%가 추천을 통해 발생한다. 아마존의 성공으로 추천 시스템은 쇼핑, 영화[1][2], 음악[3][4], 음식점[5][6] 등 다양한 분야에서 널리 활용되고 있다.

사람들의 생활방식 변화에 의해 외식 및 배달음식의 소비가 증가하고 있으며 사용자의 취향 및 상태에 따라 매년 적절한 음식을 선택하는 것은 매우 어렵다. 따라서 다양한 음식점 추천 서비스가 존재한다. 하지만 기존의 음식점 추천 서비스는 음식점의 매출향상에 초점을 맞추고 있어 사용자를 고려한 추천 방법이 요구된다. 이를 위해서는 음식점의 평가 정보와 더불어 사용자의 취향 및 상황과 같은 다양한 추가 정보가 필요하지만 대부분의 음식점 추천 서비스는 법적, 정책적 방침에 따라 데이터를 공개하지 않는다. 따라서 공개된 데이터를 기반으로 알고리즘 설계 및 성능분석을 수행해야 한다. 하지만 공개된 데이터는 알고리즘 개발에 필요한 속성을 포함하지 않기 때문에 다른 데이터의 속성 활용방법이 요구된다[7].

이와 같은 문제를 해결하기 위해 본 논문은 머신러닝 기법을 활용한 음식점 추천시스템을 제안한다. 제안하는 시스템은 자제가 기법으로 구성되며 이를 기반으로 추천 리스트를 생성한다. 첫번째, 음식점 특징 추출 기법은 자연어 문장으로 구성된 음식점 평가 의견의 특징 단어를 추출하고 이를 수치화하여 음식점 간 유사도를 계산한다. 두번째, K-means 기반의 평가 점수 유사그룹 생성 기법은 맛, 환경, 서비스 점수를 이용하여 추천, 보통, 미추천 그룹을 생성한다. 마지막 선형회귀 기법을 통해 상황 별 주문량을 계산하여 선호도를 예측한다.

제안하는 기법의 설명을 위해 본 논문은 다음과 같이 구성된다. 2장에서 추천시스템과 머신러닝 기법을 소개하고, 음식점 분야의 추천시스템 연구 동향을 분석한다. 3장에서는 머신러닝 기반 음식점 추천 메커니즘을 제안한다. 4장에서는 제안하는 시스템의 개발환경 및 추천 리스트 생성결과를 보이고 주문량 예측 기법의 성능평가를 제시하며, 5장에서 결론을 맺는다.

II. 관련 연구

2-1 추천시스템 기법

추천 기법은 사용자의 평가, 댓글, 이용내역 등을 분석하여 관심사를 파악하고 선호도가 높은 만한 아이템을 추천하는 기술이다[8]. 대표적으로 협업 필터링 기법[9]과 콘텐츠 기반 추천 기법 [10]으로 구분할 수 있다.

1) 사용자 기반 협업 필터링 기법

추천 대상과 다른 사용자의 유사도를 측정하여 유사 사용자가 높은 평가를 매긴 아이템을 추천하는 기법이다[11]. 예를 들어 대상과 비슷한 사용자b의 아이템 이용 내역에서 특정 아이템을 대상 고객에게 추천한다.

2) 아이템 기반 협업 필터링 기법

추천 대상과 다른 아이템의 유사도를 측정하여 사용자가 선호하는 아이템 대해 높게 평가한 사용자에게 추천 대상을 추천하는 기법이다[12]. 이는 데이터가 충분한 경우 정확도가 높지만 신규 사용자나 아이템과 같이 충분한 데이터가 없는 경우 Cold-Start 문제가 발생한다.

3) 콘텐츠 기반 추천 기법

콘텐츠 기반 필터링은 항목 자체를 분석하여 추천을 구현한다[13]. 이를 위해서는 항목을 분석한 프로파일과 사용자의 선호도 프로파일의 유사성을 계산한다. 이를 위해 군집분석, 인공 신경망, TF-IDF(Term Frequency Inverse Document Frequency) 등의 기술이 사용된다. 콘텐츠 기반 필터링은 Cold-Start 문제가 해결 되지만 다양한 아이템을 동시에 추천해야 하는 경우 각각의 항목 프로파일 구성이 어렵다.

2-2 머신러닝 기법

머신러닝(Machine Learning)은 명시적으로 프로그래밍 되지 않고 학습을 통해 예측이 더 정확해질 수 있도록 하는 기법이며 대표적으로 아래 기법들이 추천시스템에 활용된다.

1) 텍스트 분석 기법

컴퓨터를 통해 텍스트 분석을 수행하기 위해서는 문장, 단어를 컴퓨터가 연산 가능한 형태로 변환하는 임베딩이 요구된다. Word2Vec는 주변 단어를 이용하여 대상 단어를 예측하는 CBOW (Continuous Bag-of-Word)와 대상 단어로 주변 단어를 예측하는 Skip-gram를 지원하며 [14], 단어의 의미와 문장의 맥락을 고려하여 단어를 벡터로 표현한다.

2) K-means 알고리즘

k-Means 알고리즘[15]은 주어진 n개의 데이터를 k개로 분류하는 알고리즘이다. 원하는 목표치에 이를 때까지 계산을 반복하여 각 분류 간의 거리 차이를 최소화한다. k-means는 비교적 간단하며 빠르고 대용량 데이터에도 잘 작동한다. 하지만 클러스터의 개수(k)를 지정해야만 하며 이에 따라 도출되는 결과의 활용 방안 및 성능 편차가 크다[16].

3) 선형회귀 (Linear Regression) 기법

선형회귀[17]는 선형으로 예측 모델을 구현하는 알고리즘이다. 이 기법은 $y = ax + b$ 형태의 선형 모델을 입력 데이터를 활용하여 구축하고 이를 기반으로 종속변수를 예측한다. 이는 데이터양과 상관없이 사용이 가능하며 금융, 온라인 쇼핑 등 다양한 분야에 응용된다.

2-3 음식점 추천 연구동향

기존의 음식점 추천 시스템은 사용자의 취향과 음식점 간의 유사도를 주로 활용하였다. 주로 유클리디안 (Euclidean), 피어

슨(Pearson), 코사인(Cosine) 기법으로 유사도가 높은 음식점을 추천한다. [18]은 평가 데이터를 기반으로 8가지 특징을 도출하여 음식점을 추천하는 콘텐츠 기반 추천 시스템을 제안하였다. Nave Bayes, Decision Tree 등 다양한 분류기를 통해 만족 예측 실험을 수행하였다. 하지만 실험자가 평가 문장을 매번 수동으로 3등급으로 분류하고 8개 항목이 고정적으로 설정되어, 데이터 마다 가지고 있는 특성을 자동적으로 적용하기는 어렵다.

[19]는 Nave Bayes을 기반으로 추천 여부를 예측하는 협업 필터링 추천시스템을 제안하였다. SentiWordNet 라이브러리를 활용하여 평가 의견의 긍정적 정도를 수치화 하였으며, 성별, 나이, 음식취향, 출신, 선호 가격대, 결혼여부 등과 같은 사용자 정보를 기반으로 선호 음식점을 예측한다. 하지만 상황에 따른 취향변화나 목적에 맞는 추천을 제공하지 못하며, 식당이나 사용자 정보가 적은 콜드스타트 문제가 발생할 수 있다.

[20]은 평가점수 및 사용자 특징을 기반으로 하는 사용자 협업 필터링 기반의 추천시스템을 개발하였다. 유사도는 피어슨 공식을 이용하여 계산하며 상위 N 명의 유사사용자를 추출하여 각 사용자가 평가한 점수를 기반으로 식당을 추천한다. 이 방식 또한 특정사용자의 취향을 고정적으로 간주하기 때문에 상황 및 취향에 맞는 다양한 추천을 제공하지는 못한다.

또한, 음식점을 선택할 때 취향뿐만 아니라 시간대, 날씨 등의 상황도 영향을 준다. [5]는 모바일 환경에서 사용자의 상황을 고려하는 상황인식 추천 맛집 서비스를 제안하였다. 이는 사용자의 위치와 시간을 수집하고 맛집 데이터 시스템을 통해 교통수단에 따른 정보를 선별한다. 그 후 시간별 선호음식, 사용자 선호음식, 계절별 제철음식 등의 정보로 음식점을 추천한다.

[6]은 날씨, 온도에 따라 사용자의 데이터를 분석하여 상황에 맞는 음식점을 추천한다. [6]의 시스템은 모바일 장치로부터 위치, 날씨, 온도데이터를 수집하고 방문한 음식점의 정보, 방문 기록, 방문 후 평가 등의 데이터를 수집한다. 이후 Naive Bayes 방정식을 이용하여 사용자 행동 패턴을 분석하여 추천하고 위치, 평가 점수 등의 기준을 통해 음식점을 추천한다.

이와 같이 최근에는 취향 및 음식의 유사도 뿐 아니라 날씨, 위치, 교통정보 등의 상황정보를 추천에 활용하는 연구[5],[6]가 활발히 진행되고 있다. 또한 Cold Start와 같은 추천시스템의 문제를 해결하기 위해 다양한 기법을 활용하는 하이브리드 방식[1]의 연구가 많이 진행되고 있다. 따라서 본 논문에서는 추천 시스템 문제점을 해결하고 다양한 목적의 추천을 제공하기 위해 텍스트 분석 및 머신러닝 기법[15],[17]을 활용한 머신러닝 기반 하이브리드 방식의 추천 시스템을 설계 및 제안한다.

III. 제안하는 머신러닝기반 음식점 추천시스템

일반적으로 추천시스템은 다수의 속성을 포함하고 있는 상세한 데이터 일수록 각각의 속성을 활용한 다양한 기법을 적용할 수 있으며, 충분한 분량의 데이터가 확보되어야 안정적이며

신뢰적인 성능을 보인다. 따라서 본 논문에서는 선행 연구인 데이터 속성 확장 메커니즘 [7]을 활용하여 확장된 데이터를 기반으로 충분한 분량의 상세한 데이터를 확보하고 3가지 머신러닝 기법을 적용한 음식점 추천시스템을 제안한다.

3-1 데이터 확장 메커니즘

그림 1은 [7]에서 제안한 데이터 확장 메커니즘의 개념을 보여준다. 데이터를 확장하는 방법은 두 가지로 나뉜다.

- (1) 주 데이터셋의 특정 대상과 타 데이터셋의 동일여부를 판단할 수 있는 경우, 주 데이터셋에 존재하지 않는 속성을 외부 데이터셋에 가져와 확장하는 방법
- (2) 대상의 동일여부를 판단하지 못한 경우, 대상을 유사한 특성을 가진 대상으로 그룹화 하고 같은 그룹끼리 비교하여 주 데이터셋에 포함되지 않은 속성 값을 외부 데이터셋에서 그룹단위로 확장 하는 방법

본 논문에서 제안하는 시스템은 (1)의 기법으로 동일 음식점으로 판단되는 데이터의 평가 정보를 확장하고, (2)의 기법으로 날짜별 최저, 최고 온도 속성을 포함하는 데이터를 이용하여 특정 지역에 존재하는 음식점 그룹에 환경 속성을 확장하였다.

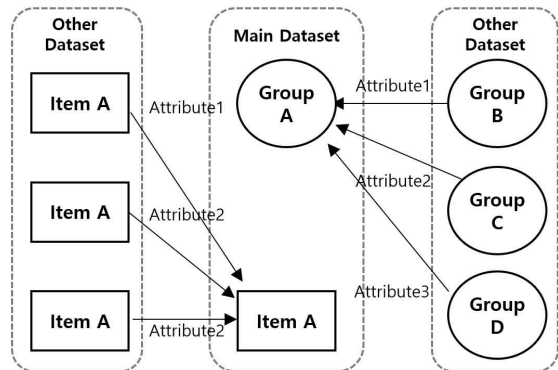


그림 1. 데이터 속성 확장 메커니즘
Fig. 1. Data Feature Extension mechanism

3-2 머신러닝 기반 음식점 추천 메커니즘

그림 2는 제안하는 머신러닝 기반 음식점 추천 메커니즘이다. 이는 3-1절에서 제안된 데이터 속성 확장 기법을 이용한 데이터를 활용하여 머신러닝의 3가지 기법으로 생성된 그룹을 통해 최종적으로 사용자에게 음식점을 추천한다.

첫 번째는 자연어 평가 기반 음식점 특징 추출 기법으로 자연어 문장으로 된 음식점 평가 데이터에서 특징을 추출하여 유사그룹을 생성한다. 두 번째는 K-means 기반의 평가 점수유사 그룹 생성 기법으로 사용자들의 평가 점수를 그룹화 하여 유사한 평가를 받은 그룹을 생성한다. 마지막은 선형회귀 기반 상황별 음식점 주문량 예측 기법으로 날씨, 계절 등의 상황정보 및 주문량 정보를 기반으로 현재 상황에서의 주문량을 예측한다.

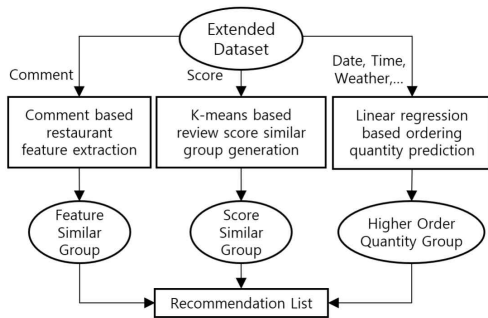


그림 2. 제안하는 머신러닝 기반 음식점 추천 메커니즘
 Fig. 2. Proposed machine learning based restaurant recommendation mechanism

1) 자연어 평가 기반 음식점 특징 추출 기법

자연어 평가 기반 음식점 특징 추출 기법은 음식점 특징을 추출하기 위해 그림 3과 같은 단계를 거친다.

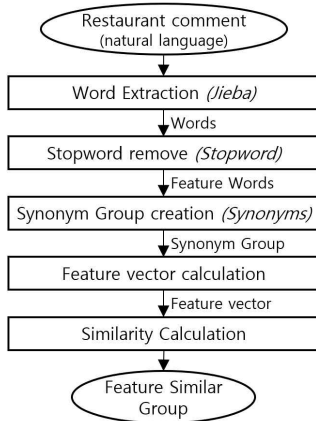


그림 3. 자연어 평가 기반 특징 유사그룹 생성기법
 Fig. 3. Natural language comment based feature similar group creation method

먼저 자연어 문장으로 되어있는 음식점 평가의견을 입력한다. 본 논문에서는 북경 대학교 개방형 연구 데이터 플랫폼에서 제공하는 Dianping.com의 음식점 평가 데이터[21]를 활용하였다. 그 이유는 해당 데이터가 자연어로 구성된 다양한 평가의견을 포함하고 있으면서 CC0 - "Public Domain Dedication" 라이선스로 배포되어 저작권에 자유로운 데이터이기 때문이다. Dianping.com은 중국 사용자에게 상점정보, 소비자 평가 등 생활 서비스를 제공하는 O2O 플랫폼으로 활용 한 데이터는 중국어 문장으로 구성되어 있다. 중국어 문장 처리를 위해 중국어 문장의 단어 추출 기능을 제공하는 jieba 라이브러리로 문장에서 단어를 추출한다. 추출된 모든 단어는 명사, 동사, 조사 등 다양한 항목들이 포함될 수 있다. 이 중 특징을 나타낼 수 있는 단어만을 추출하기 위해 Stopwords를 사용한다. 이는 중국어 불용어 리스트로 여기에 포함된 단어를 제거한다. 이후 synonyms 라이브러리로 유의어 그룹을 생성한다. 해당 라이브러리는 특정 단어를 기준으로 유사도를 계산하여 동의어 및 유의어를 찾

도록 한다. 생성된 그룹은 해당 음식점에 남겨진 다양한 평가의견의 대표적인 단어 그룹으로써 해당 음식점의 특징이 된다. 본 논문에서는 8개의 유의어 그룹을 생성하여 8개의 특징 벡터를 계산한다. 식 (2)는 특징 그룹에 포함된 단어의 출현 횟수를 기반으로 특징 벡터 값을 계산한다. 값이 클수록 해당 그룹에 속한 단어의 발생 빈도가 높으며, 이는 해당 음식점에 대한 평가의견에서 빈번하게 발생되는 단어라고 할 수 있다. 모든 유의어 그룹의 발생빈도 계산을 통해 특징 벡터 값을 구한다.

$$feature V_n = \frac{count_n}{count_1 + \dots + count_8} \quad (2)$$

결과적으로 음식점당 8가지 특징 벡터 값을 가지며 이를 기반으로 유사한 특징을 가진 음식점의 목록을 구할 수 있다.

2) K-means 기반의 평가점수 유사그룹 생성기법

K-means 알고리즘은 다양한 분포를 가지고 있는 대상을 k개의 그룹으로 나누어 주는 머신러닝 알고리즘이다. 본 논문은 이를 기반으로 유사한 평점을 가진 음식점 그룹을 생성한다. 그림 4는 제안하는 k-means 기반 평가점수 유사그룹 생성기법이다.

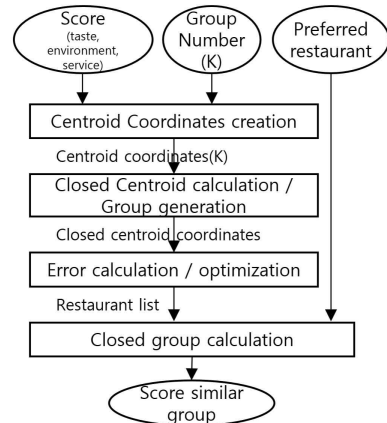


그림 4. 평가점수 유사그룹 생성기법
 Fig. 4. Review score similar group creation method

음식점의 평점은 맛, 환경, 서비스 점수로 구성되며 0-5까지 값을 가진다. 유사그룹 생성을 위해 먼저 클러스터 개수를 결정 (K)하고 임의의 중심(Centroid) 좌표를 생성한다. 이때, 입력되는 평점이 3가지 항목을 가지고 있으므로, 3차원 좌표를 가진다. 입력된 모든 평점은 자신과 가장 근접한 중심을 찾고 해당 중심을 기반으로 그룹을 설정한다. 이와 같이 설정된 그룹을 기준으로 평균을 계산하고 이를 새로운 중심으로 설정한다. 보정된 중심을 기준으로 최적화(중심점 좌표가 변경되지 않거나, 오차범위 이내로 설정된 경우) 할 때까지 반복한다. 그룹 구성이 최적화 되면, 사용자로부터 기준이 되는 선호 음식점의 평점 정보를 받아 해당 음식점의 평점이 포함되는 그룹을 찾는다. 결과적으로 도출된 그룹은 사용자가 선호하는 음식점과 유사한 평점을 가진 음식점의 목록을 포함하고 있다. 이때 n의 크기가 너무 크면 너무 적은수의 구성원을 갖는 다수의 그룹으로 분할

되며, n의 크기가 너무 작으면 다수의 멤버가 하나의 그룹으로 분할되어 상세한 취향에 맞는 리스트를 생성하기 어렵다.

3) 선형회귀 기반 상황별 음식점 주문량 예측 기법

음식점의 주문량은 날씨, 이벤트, 계절 등의 상황에 따라 변동되며 주문량 예측을 통해 재고 관리 및 해당 상황의 선호도를 예측하여 추천할 수 있다. 식 (3)은 음식점 주문량을 예측 위한 선형회귀 모델이다. 상황 정보는 선형회귀 모델의 입력인 독립변수이고, 음식점 주문량의 예측 값은 종속변수이다. $H(x)$ 는 종속변수인 주문량 예측 값이고, w_1-w_5 는 선형회귀 모델링된 독립변수별 가중치이다. 독립변수인 X_{LT} 는 최저온도, X_{HT} 는 최고온도, X_W 는 날씨, X_M 는 월, X_D 는 날짜, e 는 오차 항이다. 축적된 날씨 및 주문량 데이터를 활용하여 추천 상황의 주문량을 예측한다. 예측된 주문량이 높다면 현재 해당 음식점의 선호도가 높을 것으로 간주하고 주문량 상위 음식점 그룹을 생성한다.

$$H(x) = w_1 X_{LT} + w_2 X_{HT} + w_3 X_W + w_4 X_M + w_5 X_D + e \quad (3)$$

IV. 음식점 추천시스템의 실험 및 성능 분석

4장에서는 제안하는 음식점 추천시스템의 실험 및 성능 분석을 수행한다. 먼저 4.1장에서는 추천시스템 실험을 위한 데이터 셋과 추천시스템 개발 환경에 대해 소개하고 4.2장에서는 제안하는 추천시스템의 3가지 기법 (자연어 평가기반 특징추출, k-means 기반 평가점수 유사그룹, 선형회귀 기반 상황별 주문량 예측)을 통한 실험 결과를 제시한다. 마지막으로 4.3장에서는 추천 결과의 정확성을 판단하기 위한 성능 평가를 수행한다.

4-1 제안하는 음식점 추천 시스템 실험 및 개발 환경

1) 음식점 데이터 셋

표 1은 제안하는 시스템의 성능 분석에 사용된 데이터 형식이다. 앞서 설명한 음식점 평가 데이터[21]는 206개 음식점에 대한 평가정보 453,848건을 포함하고 있다. 사용자 관련 정보는 사용자 아이디, 사용자 등급 등을 포함하고 있으며, 음식점 관련정보는 음식점 아이디, 지출액 등이 있다. 또한 평가 정보는 총점, 맛 점수, 환경 점수, 서비스 점수, 평가 시간 및 평가 의견을 포함하고 있으며 점수의 경우 0~5의 값을 갖는다.

표 1. 추천 실험 데이터 셋 형식

Table 1. Recommend experiment dataset format

Name	Description
Review_ID	Review ID
Merchant	Restaurant name
Rating	Restaurant rating
Score_taste	Taste rating
Score_environment	Environmental rating
Score_service	Service rating

Price_per_person	Per capita price (Null means empty)
Time	Commenting time
Num_thumbs_up	Like count
Num_response	Reply count
Content_review	Comment text
Reviewer	Commenter's user name
Reviewer_value	Commenter's level
Reviewer_rank	If commenter is vip
Favorite_foods	Favorite foods

2) 음식점 추천 시스템 개발 환경

본 절은 제안하는 추천 시스템의 개발 환경을 설명한다. 개발 환경은 표 2와 같이 Window 10 환경에서 python 3.7을 사용하였다. 개발 도구는 Anaconda, Tensorflow, Jupyter를 사용하였다. Jupyter는 Web기반 응용프로그램으로 데이터 처리와 변환, 수치 시뮬레이션, 통계 모델링, 기계 학습 등을 제공한다. 또한 pandas, numpy 라이브러리를 이용하여 음식점 데이터를 처리 및 분석하고 jieba를 이용하여 중국어 자연어 처리 및 특징 단어를 추출한다. 또한 synonyms를 이용하여 동의어, 및 유의어를 추출하고 그래프 출력을 위해 matplotlib을 사용한다.

표 2. 시스템 개발환경

Table 2. System implementation environment

Environment	Name	Purpose
OS	Win10	
Programming Language	Python 3.7	Machine Learning and data processing
Development Tools	Anaconda	Conda package managing
	Jupyter	Development tool
Libraries	Pandas	Data pre-processing
	Tensorflow	Machine Learning
	Numpy	Number processing
	Jieba	Chinese language processing
	Matplotlib	Visualization
	Synonyms	Chinese synonyms extraction
Dataset	Dianping Dataset	www.dianping.com

4-2 음식점 추천 실험 결과

1) 자연어 평가 기반 음식점 특징 추출 결과

표 3은 synonyms 라이브러리를 이용하여 유의어 그룹을 생성한 결과이다. 표의 첫 번째 열은 도출된 8가지 특징 단어이며 이를 기준으로 유의어 도출결과를 보여준다. 예를 들어 여행과 같은 의미로 추출된 단어는 여행 관광, 관광, 투어 서비스, 관광 여행, 투어, 여행자, 상태 관광, 쇼핑, 관광지역이 있다.

표 3. synonyms 기반 유의어 그룹 생성 결과

Table 3. Synonyms based synonym group creation result

Feature Word	Synonyms
여행	관광, 투어 서비스, 관광여행, 투어, 여행자, 생태, 관광, 쇼핑, 관광, 지역
맛집	과자, 요리, 디저트, 요리사, 조리, 맛, 해산물, 소식,

	별미
모임	부도, 회식, 예배, 파티, 연회, 설교, 예배하기, 미팅, 친한친구
교통	교통운송, 도로교통, 교통시설, 도시교통, 대중교통, 철도운송, 운송, 교통네트워크, 교통망
커플	연인, 부부, 데이트, 여자친구, 남자친구, 부녀, 2명, 전 여친, 두분
건강	몸이 건강, 심리건강, 보건, 심신건강, 위생, 생활품질, 비만, 비만증, 유익함
영양	탄수화물, 영양소, 음식, 양분, 영양성분, 영양물질, 미량원소, 섭취, 당류
간식	별식, 분식, 두부, 샤브샤브, 쇠고기면, 떡, 사천요리, 반찬, 과자

표 4는 206개 음식점 별로 특징단어그룹의 발생빈도를 기반으로 특징벡터 값을 계산한 결과이다. 표 4의 세로는 8가지 특징단어를 나타내며, 가로 R1 ~ R206은 음식점의 ID를 나타낸다. 한 음식점의 모든 특징단어 벡터 값의 합은 1이 되며 206개 음식점마다 계산된 결과는 특징 벡터 행렬로 저장된다.

표 4. 음식점 특징벡터 계산 결과

Table 4. Restaurant feature vector calculation result

	R1	R2	...	R205	R206
여행	0	0.01221	...	0.059322	0.003878
맛집	0.51095	0.4579	...	0.381356	0.413184
모임	0.07664	0.02893	...	0.050847	0.046532
교통	0.00365	0.02078	...	0.169492	0.048686
커플	0.0073	0.00686	...	0.076271	0.003447
건강	0.0438	0.01521	...	0.084746	0.019819
영양	0.05839	0.04907	...	0.059322	0.075399
간식	0.29197	0.40583	...	0.09322	0.383455

표 5는 피어슨 유사도 기법[22]을 이용한 206개 음식점간 유사도 계산 결과이다. 유사도는 0~1의 값을 가지며 특정 음식점을 기준으로 206번의 계산이 요구된다. 예를 들어, 사용자가 R1을 선호하는 경우, R1과 모든 음식점 간의 유사도를 계산하여 사용자에게 추천한다. 유사도가 높으면 사용자가 선호하는 음식점과 특징이 매우 비슷하다고 생각할 수 있다. 반대로 유사도가 낮으면 선호 음식점과 특징이 매우 다르기 때문에 그 음식점을 선호하지 않을 수 있다.

표 5. 206개 음식점 유사도 계산 결과

Table 5. Similarity calculation result for 206 restaurants

	R1	R2	...	R99	R100	...	R204	R206
R1	1.0000	0.9644	...	0.8966	0.9184	...	0.9661	0.9569
R2	0.9644	1.0000	...	0.9271	0.8346	...	0.9906	0.9962
...
R99	0.8966	0.9271	...	1.0000	0.8764	...	0.9348	0.9457
R100	0.9184	0.8346	...	0.8764	1.0000	...	0.8590	0.8484
...
R204	0.9661	0.9906	...	0.9348	0.8590	...	1.0000	0.9911
R206	0.9569	0.9962	...	0.9457	0.8484	...	0.9911	1.0000

2) K-means 기반의 평가점수 유사그룹 생성결과

실험 데이터 셋은 453,848개의 평가 데이터를 포함하며 음식점별로 평가 점수의 평균을 계산하여 k-means의 입력 데이터를 생성하였다. 하나의 음식점은 3개의 좌표 값을 가지는 하나의 점으로 표현되며 x 좌표는 맛, y좌표는 환경, z는 서비스 점수를 나타낸다. 그룹의 개수를 설정하는 중심점의 개수 K는 각 점수를 상, 중, 하로 구분한다고 가정하여 3ⁿ으로 설정하였다. 평가 점수를 3가지로 구분 하므로 n을 1,2,3으로 설정하였다.

표 6은 K=3일 때 (그룹의 개수가 3개일 때) 그룹을 구성하는 클러스터 중심점의 계산 결과이다. 206개의 음식점 점수 데이터를 기반으로 3개의 중심점을 이용하여 가장 가까운 중심점을 계산하면 해당 그룹의 음식점 리스트 3개를 생성할 수 있다. 예를 들어, G1 음식점 리스트의 기준은 맛 점수가 3.1, 환경 점수가 3.2, 서비스 점수가 3.2이다.

표 6. K=3일 때, 클러스터 중심점

Table 6. Cluster centroid coordinates (if K=3)

	G1	G2	G3
x (taste)	3.1	3.5	3.3
y (environment)	3.2	3.7	3.1
z (service)	3.2	3.7	3.1

그림 5는 K=3일 때 그룹화 결과 그래프이다. 그래프의 x 좌표는 맛, y 좌표는 환경, z 좌표는 서비스 점수이다. 평가 점수는 0~5 사이의 값을 가지고 있지만 입력으로 사용된 음식점 평균 평가 점수는 2~4 사이의 값만 존재하여 데이터 확인을 보다 용이하게 하도록 그림 5의 그래프는 2~4점 구간을 표현하도록 설정하였다. 그림 5에서 파랑색 삼각형으로 표시되는 음식점들은 3가지 평가 점수가 모두 높은 음식점 그룹이다. 이 그룹은 사용자에게 적극 추천할 수 있다. 중간 영역에 존재하는 녹색 사각형들은 3가지 점수가 모두 보통인 그룹이다. 이 그룹에 속하는 음식점은 사용자에게 보통의 순위로 추천할 수 있다. 빨강색 원형으로 표시되는 음식점들은 점수가 낮은 음식점들이 그룹화된 것이다. 이 그룹의 음식점은 대체적으로 추천하지 않는다. 이와 같이 평가 점수에 따라 3가지 그룹으로 음식점을 분류하여 사용자에게 적극추천/보통/비추천 목록을 제공할 수 있다.

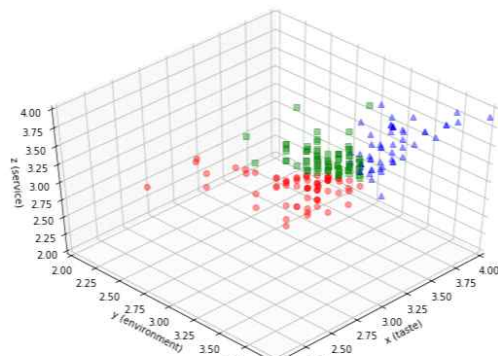


그림 5. K=3일 때, K-means 기반 그룹화 결과
Fig. 5. K-means based grouping result (if K=3)

표 7은 K=9일 때 (그룹의 개수가 9개 일때) 클러스터의 중심점 계산 결과이다. 이를 기반으로 9개의 그룹으로 구분된 리스트를 생성할 수 있다. 예를 들어, G1 음식점 리스트의 기준은 맛 점수 3.8, 환경 점수가 3.5, 서비스 점수가 3.5이다.

표 7. K=9일 때, 클러스터 중심점
Table 7. Cluster centroid coordinates (if K=9)

	G1	G2	G3	G4	G5	G6	G7	G8	G9
x	3.8	3.2	3.4	3.1	3.3	3.2	3.3	3.1	3.6
y	3.5	2.5	3.3	3.4	3.4	3.1	3.3	2.9	3.6
z	3.5	2.5	3.3	3.4	3.4	3.1	3.3	2.9	3.6

그림 6은 K=9일 때 그룹화 결과 그래프이다. K를 3으로 설정하여 그룹화한 그림 5에 비해 K가 큰 값을 가지기 때문에 많은 그룹으로 분류된 것을 확인할 수 있다. 그림 6을 보면 빨간 동그라미와 파란 마름모로 표현된 두 개의 그룹이 모두 적극 추천할 만한 높은 점수를 가지고 있는 것을 확인할 수 있으며 그림 5에 비해 세부적으로 분류 되어 사용자의 취향을 세분화 하여 만족 시킬 수 있을 것으로 생각된다. 예를 들어 사용자 A는 환경점수가 다소 낮더라도 맛과 서비스 점수가 높은 경우를 선호하는 반면, 사용자 B는 환경 및 서비스 점수를 우선적으로 생각할 수 있다. 따라서 평균적으로는 두 사용자가 선호하는 음식점의 점수가 상위권에 포함되면서도 세부적인 점수의 차이가 취향의 차이로 대표될 수 있다. 하지만 무분별하게 그룹의 수를 증가시키면 특정 그룹에 충분히 많은 데이터가 포함되지 못해 의미 있는 결과를 도출하지 못할 수 있다.

그림 7은 K=27일 때 그룹화 결과 그래프이다. 모든 대상이 그림 6에 비해 매우 세부적으로 분류되었다. 하지만 앞서 설명한 대로 너무 많은 그룹으로 분류할 경우 그룹별 표본수 매우 적어 의미 있는 활용이 어렵다. 또한 입력 데이터인 평가 점수가 주관적인 선호도를 표현하기 때문에 그림 7과 같이 매우 세부적인 분류는 필요하지 않을 것으로 생각된다. 따라서 평가점수 기반의 음식점 그룹화는 K가 3일 때, 추천/보통/비추천과 같은 확실한 목적에 맞는 경우와 K가 9일 때, 세부적인 취향을 적용하는 정도가 적절할 것으로 생각된다.

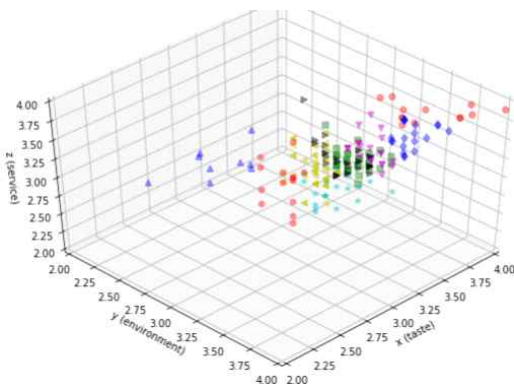


그림 6. K=9일 때, K-means 기반 그룹화 결과
Fig. 6. K-means based grouping result (if K=9)

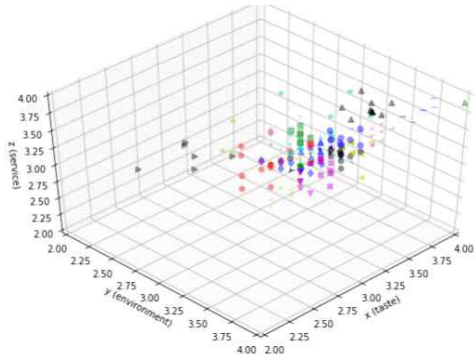


그림 7. K=27일 때, K-means 기반 그룹화 결과
Fig. 7. K-means based grouping result (if K=27)

3) 선형회귀 기반 상황별 음식점 주문량 예측 결과

표 8은 음식점 월 주문량 예측 결과이다. 2014년부터 2016년까지 월별 주문량, 월별 평균 온도, 월 정보를 이용하여 2019년 월별 주문량을 예측하였다. 첫 행은 월을 나타내며 첫 열은 음식점 아이디이다. 또한 일별 데이터를 입력하면 특정 날짜의 주문량을 예측할 수 있다. 예측된 주문량이 많을수록 특정 상황(날씨, 월, 일 등)에서 해당 음식점의 선호도(인기)가 많을 것으로 간주하고 추천 할 수 있다.

표 8. 음식점 주문량 예측 결과
Table 8. Ordering quantity prediction result

	1	2	3	4	5	6	7	8	9	10	11	12
R1	74	49	27	32	28	15	16	16	16	14	8	12
R2	856	719	605	497	444	449	505	394	262	326	260	244
R3	427	437	346	284	213	240	314	284	156	182	193	158
...
R99	125	114	92	44	56	44	33	49	22	27	15	23
R100	134	107	120	101	112	78	99	85	55	47	36	43
R101	5	6	50	52	75	58	77	60	46	60	35	37
...
R204	133	195	98	76	90	89	120	118	112	129	114	123
R205	3	6	7	7	13	16	56	33	28	49	51	55
R206	290	182	213	144	155	130	149	123	66	65	48	54

4) 추천 리스트 생성 결과

본 절에서는 제한한 3가지 기법을 통한 음식점 추천리스트 생성 결과를 보여준다. 표 9는 자연어 평가 기반 음식점 특징 추출 기법으로 생성된 음식점 추천 리스트이다. R1을 기준으로 유사도를 계산했기 때문에 R1과의 유사도는 1로 계산된 것을 확인할 수 있다. 유사도가 높은 음식점 순서대로 유사도와 평점을 제공한다. 하지만 유사도가 높다고 해서 가장 높은 평점을 가지고 있지는 않다. 이 리스트는 특정 음식점과 유사한 평가를 받은(특징이 유사한) 음식점 순위일 뿐이지 무조건 높은 평가를 받은 순위가 아니기 때문이다. 특정 사용자가 R1을 선호하는 경우 R1 음식점과 유사도 높은 음식점을 추출하여 사용자에게

게 추천한다. 사용자는 이를 기반으로 자신이 선호하는 음식점과 유사한 평가를 받은 다른 음식점을 이용할 수 있다.

표 9. 평가 기반 추천 리스트

Table 9. Comment based recommendation list

Rank	Restaurant_ID	Similarity	Rating
1	R1	1.00000	4.3
2	R27	0.99566	4.3
3	R181	0.99335	4
4	R140	0.99285	4.4
5	R109	0.99209	4.3
6	R182	0.98958	4.2
7	R28	0.98816	4.1
8	R172	0.98793	4.4
9	R29	0.98624	4.2
10	R189	0.98419	4.3

표 10은 K-means 기반으로 도출한 유사 평가점수 음식점 추천 리스트이다. 예를 들어, 특정 사용자가 R21을 선호하면 해당 음식점의 맛, 환경, 서비스 점수와 유사한 점수를 받은 음식점의 목록을 제공하며 이를 통해 해당 평가 항목을 중요시하는 사용자의 취향에 맞는 음식점을 선택 할 수 있다. 표를 보면 세부적인 평가 점수와 종합 평가 점수의 차이가 있는 것을 확인할 수 있다. 기준이 되는 R21과 유사도가 가장 높은 R121의 경우 종합점수가 4.5점인 반면, 상대적으로 유사도가 약간 낮은 R29의 경우 4.2 점으로 기준 음식점인 R21의 5점에 비해 종합점수의 차이가 발생하나 세부적 평가항목인 맛, 환경, 서비스 점수의 유사도는 높은 것으로 확인되었다.

표 10. 평가점수 기반 추천 리스트

Table 10. Review score based recommendation list

Rank	Restaurant_ID	Similarity	Rating
1	R21	1	5
2	R121	0.999999	4.5
3	R156	0.999998	4.4
4	R187	0.999998	4.4
5	R115	0.999997	4.3
6	R124	0.999997	4.3
7	R29	0.999995	4.2
8	R46	0.999929	4.5
9	R58	0.999929	4.8
10	R59	0.999929	4.8

표 11은 선형회귀기법으로 예측한 주문량을 기반으로 생성된 추천 리스트이다. 결과를 보면 주문량이 가장 높은 R23의 종합점수는 4.1이다. 하지만 주문량 10위인 R113의 경우 더 높은 4.7점이다. 이는 현재 주문량이 많다고 해서 종합적인 평가점수가 높은 것이 아님을 알 수 있다. 날씨, 날씨와 같은 상황에 따른 주문량은 해당 시점에 인기 있는 음식점으로 간주 할 수 있으며, 이를 기반으로 현재 인기 있는 음식점을 추천할 수 있다.

표 11. 주문량 기반 추천 리스트

Table 11. Order quantity based recommendation list

Rank	Restaurant_ID	count	Rating
1	R23	583	4.1
2	R19	358	4.4
3	R2	326	4.4
4	R86	244	3.7
5	R106	244	4
6	R8	241	4.3
7	R148	241	4
8	R132	205	4.1
9	R128	204	4.7
10	R113	197	4.7

4-3 음식점 추천 성능 평가

추천시스템의 성능분석은 추천을 받은 대상의 만족도 데이터가 있는 경우 수행할 수 있지만 해당 데이터가 없는 경우 성능분석이 어렵다. 따라서 본 절에서는 성능 분석이 가능한 항목인 주문량 예측 결과에 대한 성능분석을 수행한다. 예측 주문량의 정확성을 평가하기 위해 실제 주문량과 예측 주문량의 MAE(Mean Absolute Error)를 계산하였다. 하지만 주문량의 범위가 정해지지 않은 상황에서는 절대적인 수치를 비교하기 어렵다. 따라서 그림 8과 같이 주문량 대비 오차 비율을 계산하였다. 전체 206개 음식점 중 오차 확률이 10%인 음식점은 132개, 20%는 22개로 나타났다. 예측 결과, 대부분이 (206개중 167개 : 81%) 오차범위 10%~30%안에 존재하여 주문량의 추세를 판단하기에 충분한 성능을 보여준다고 생각된다.

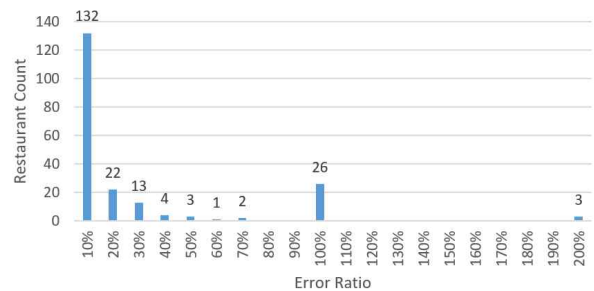


그림 8. 예측 주문량 오차 비율

Fig. 8. Error Ratio of order quantity prediction

표 12는 제안하는 방법과 기존 방법들과의 특징 비교 분석결과이다. [18],[19]는 머신러닝 기법을 통해 사용자 및 관련 데이터를 협업 필터링하여 추천여부를 결정하나 추출되는 특징의 개수 및 종류가 고정적이어서 데이터 특성에 따른 다양한 특징을 반영하기 어려운 단점을 가지고 있다. 또한 상황정보를 활용하지 않아 대상의 취향을 고정적으로 모델링 한다. [5],[6]은 식당 정보 및 사용자 상황을 중심으로 식당을 추천하나 특징 추출 기능이 없어 변화하는 추천대상의 특징을 적용하지 못한다. 반면 제안하는 방법은 자연어 데이터를 동적으로 처리하여 발생

빈도가 높은 키워드를 특징으로 추출하기 때문에 데이터에 따라 추출되는 특징의 종류 및 개수가 유동적이며, 자동적으로 처리된다. 또한 상황정보를 기반 주문량 예측 기법을 통해 상황에 따른 추천을 가능하게 한다.

표 12. 관련 추천시스템과 특징 비교

Table 12. Comparisons of relative recommendation systems

	[18]	[19]	[5]	[6]	proposed
Type	collaborative	collaborative	data analysis	data analysis	hybrid
Feature Extraction	static	static	not support	not support	dynamic
Situation Awareness	not support	not support	support	support	support

V. 결론

사용자는 수많은 상품이나 정보 중 자신이 원하는 것을 편하게 찾고자 한다. 따라서 쇼핑, 영화, 음악, 음식점과 같이 많은 분야에서 추천 시스템이 활용되고 있다. 하지만 기존의 추천 시스템은 사용자나 아이템의 유사도를 추천에 활용하였다. 그러나 실제로 사용자가 대상을 선택 할 때는 현재 위치, 사용목적, 가격, 날씨, 시간 등 다양한 상황 요소를 고려하기 때문에 이를 세밀히 분석하는 기술이 필요하다.

따라서 본 논문에서는 다양한 목적 및 상황에 맞는 추천을 제공하기 위해 머신러닝 기반의 3가지 기법을 적용한 추천시스템을 제안하였다. 첫 번째, 자연어 평가 기반 음식점 특징 추출 기법은 자연어로 구성된 음식점의 후기 데이터에서 특징 단어를 추출하고 특징 벡터를 계산하여 유사 특징 음식점을 추천한다. 두 번째, K-means 기반의 평가점수 유사그룹 생성기법은 음식점의 3가지 평가 점수(맛, 환경, 서비스)를 기반으로 유사한 평가를 받은 음식점의 그룹을 생성, 추천한다. 실험결과 클러스터 개수를 3으로 두었을 때 3가지 그룹으로 음식점을 구분하여 사용자에게 적극추천/보통/비추천 음식점 목록을 제공할 수 있었으며, 클러스터 개수가 9일 때 더 세분화된 그룹으로 음식점을 분류하여 사용자의 다양한 취향을 만족 시킬 수 있었다. 하지만 무분별하게 그룹의 수를 증가시키면 특정 그룹에 충분한 표본이 포함되지 못해 의미 있는 결과를 도출하기 어렵다. 마지막, 선형회귀 기법을 기반으로 날씨, 날짜 등 상황 별 음식점 주문량을 예측하여 현재 인기 있는 음식점을 추천한다. 하지만 주문량은 특정 시점의 인기도이기 때문에 주문량이 많다고 해서 평가 점수가 가장 높지는 않은 것으로 확인되었다. 이와 같이 제안한 3가지 기법을 사용하여 상황, 목적에 따라 사용자를 만족 시킬 수 있는 음식점 리스트를 추천할 수 있다.

또한, 제안한 시스템의 성능분석을 위해 선형회귀 기반으로 예측된 상황별 주문량의 오차 비율을 계산하였다. 206개 음식점 중 주문량 오차 비율이 20% 이하인 음식점이 155개로 주문량 예측을 통해 인기의 추세를 판단하기에 충분한 것으로 판단

된다. 결과적으로 제안하는 시스템은 영화, 여행, 의료 등 다양한 분야의 추천 서비스에 활용 될 수 있을 것으로 예상되며 향후 다양한 분야의 추천 시스템 플랫폼을 개발할 예정이다.

감사의 글

이 논문은 2017년도 정부(교육과학기술부)의 재원으로 한국연구재단의 기초연구사업 지원을 받아 수행된 것임(과제번호: NRF-2017R1D1A1B03035517)

참고문헌

[1] I. Y. Jeong, X. Yang and H. K. Jung, "A Study on Movies Recommendation System of Hybrid Filtering-Based," *Journal of the Korea Institute of Information and Communication Engineering*, Vol. 19, No. 1, pp. 113-118, January 2015.

[2] D. G. Shin, S. H. Lee and J. S. Cho, "Vector Similarity based Movie Recommend System using a Large-scale Movie Data," in *Proceeding of the Korea Information Science Society*, pp.1857-1859, December 2016.

[3] D. J. Kim and K. M. Lim, "A Music Recommendation Method Using Emotional States by Contextual Information," *Journal of the Korea Society of Computer and Information*, Vol. 20, No. 10, pp. 69-76, October 2015.

[4] J. Y. Chung and M. J. Kim, "A Study on Personalized Music Recommendation Model through Analysis on Users' Music Preference Factors," *The Journal of Digital Contents Society*, Vol. 19, No. 11, pp. 2041-2047, November 2018.

[5] J. M. Ryu, C. P. Hong, K. B. Kang, D. H. Kang, D. Y. Yang and J. W. Jwa, "Development of Mobile Context Awareness Restaurant Recommendation Services," *Journal of the Korea Contents Association*, Vol. 7, No. 5, pp. 138-145, May 2007.

[6] B. I. Ahn, K. I. Jung and H. L. Choi, "Mobile Context Based User Behavior Pattern Inference and Restaurant Recommendation Model," *The Journal of Digital Contents Society*, Vol. 18, No. 3, pp. 535-542, June 2017.

[7] Q. Peng, H. S. Choi and W. S. Rhee, "Dataset Integration and Attribute Extension Method for Restaurant Recommendation System," in *Proceeding of the KICS Conference*, Jeju, pp. 1259-1260, June 2019.

[8] J. E. Son, S. B. Kim, H. J. Kim and S. Z. Cho, "Review and Analysis of Recommender Systems," *Journal of the Korean Institute of Industrial Engineers*, Vol. 41, No. 2, pp. 185-208, April 2015.

[9] M. D. Ekstrand, J. T. Riedl and J. A. Konstan, "Collaborative Filtering Systems," *Foundations and Trends in Human-Computer Interaction*, Vol. 4, No. 2, pp. 81-173, February 2011.

[10] P. Lops, M. De Gemmis, and G. Semeraro, Content-based Recommender Systems: State of the Art and Trends, Recommender Systems handbook, *Springer*, pp. 73-105, 2011.

[11] Z. D. Zhao and M. S. Shang, "User-based Collaborative Filtering Recommendation Algorithms on Hadoop," in *2010 Third International Conference on Knowledge Discovery and Data Mining*, Phuket, pp. 478-481, January 2010.

[12] B. Sarwar, G. Karypis, J. Konstan and J. Riedl, "Item-based Collaborative Filtering Recommendation Algorithms," *Proceedings of the 10th International Conference on World Wide Web*, pp.285-295, April 2001.

[13] B. W. Seo, "The Evolution of Content Recommendation Algorithm," *Broadcasting Trend & Insight*, Vol. 5, pp. 19-24, May 2016.

[14] S. M. Kim, I. S. Na and J. H. Shin, "A Method on Associated Document Recommendation with Word Correlation Weights," *Journal of Korea Multimedia Society*, Vol. 22, No. 2, pp. 250-259, February 2019.

[15] k-means clustering [Internet]. Available: https://en.wikipedia.org/wiki/K-means_clustering.

[16] Guido and Sarah, Introduction to Machine Learning with Python, *Oreilly & Associates Inc*, May 2015.

[17] Linear regression [Internet]. Available: https://en.wikipedia.org/wiki/Linear_regression.

[18] S. Zhang, M. Salehan, A. Leung, I. Cabral and N. Aghakhani, "A Recommender System for Cultural Restaurants Based on Review Factors and Review Sentiment," in *Americas Conference on Information System*, New orleans, pp. 1-5, 2018.

[19] N.G. Bhojne, S. Deore, R. Jagtap, G. Jain and C. Kalal, "Collaborative Approach based Restaurant Recommender System using Naive Bayes," *International Journal of Advanced Research in Computer and Communication Engineering*, Vol. 6, No. 4, pp. 6- 13, April 2017.

[20] A.A. Fakhri, Z.K.A. Baizal and E.B. Setiawan, "Restaurant Recommender System Using User-Based Collaborative Filtering Approach: A Case Study at Bandung Raya Region," In *Journal of Physics: Conference Series*, Vol. 1192, No. 1, pp. 1-7, 2019.

[21] Peking University Open Research Data Platform V1, "Restaurant Review Data on Dianping.com," [Internet]. Available: <https://opendata.pku.edu.cn/dataset.xhtml?persistentId=doi:10.18170/DVN/GCIUN4>

[22] S. V. Dongen, and A. J. Enright, "Metric Distances Derived from Cosine Similarity and Pearson and Spearman Correlations," *arXiv preprint*, arXiv:1208.3145, August 2012.



최환석(Hoan-Suk Choi)

2009년 : 한밭대학교 멀티미디어공학과 (공학사)

2011년 : 한밭대학교 멀티미디어공학과 (공학석사)

2018년 : 한밭대학교 멀티미디어공학과 (공학박사)

2015년~현 재: 한국 ITU연구위원회 국제표준전문가
2018년~현 재: 한밭대학교 멀티미디어공학과 박사후연구원
※ 관심분야 : Mobility Management, IoT, Social IoT, Semantic Processing, Service Composition, Metadata modeling, Trust management



팽전(Qian Peng)

2018년 : 한밭대학교 정보통신공학과 (공학사)

2018년~현 재: 한밭대학교 멀티미디어공학과 석사과정
※ 관심분야 : 추천시스템, 머신러닝, IoT



이우섭(Woo-Seop Rhee)

1983년 : 홍익대학교 (공학사)

1995년 : 충남대학교 (공학석사)

2003년 : 충남대학교 (공학박사)

1983년~2005년: 한국전자통신연구원 팀장/책임연구원

2005년~현 재: 한밭대학교 정보통신공학과 교수

2006년~현 재: 한국ITU연구위원회 국제표준전문가

2012년~2013년: Institute TelecomSudParis 방문교수

2018년~2019년: Liverpool John Moores University 방문교수

※ 관심분야 : Semantic Processing, Trust Management, IoT, Social IoT