숙소의 적정 가격 결정을 위한 Web Site 개발

조규철*, 노현진O, 송우현**
OO인하공업전문대학 컴퓨터정보과,
*인하공업전문대학 컴퓨터정보과,
**인하대학교 전기컴퓨터공학과
e-mail: kccho@inhatc.ac.kr*, rhj0622@gmail.comO, suh6911@naver.com**

Developing Web Site for Setting a Price of Accommodation

Cho Kyu Cheol*, Roh Hyun JinO, Song Woo Hyeon**
ODept. of Computer Science, Inha Technical College,
*Dept. of Computer Science, Inha Technical College,
**Dept. of Electrical Engineering and Computer Science, Inha University

● 요 약 ●

호스트가 숙소 가격을 정할 때, 기존 숙박 플랫폼들이 제공하는 최적화된 가격을 참고하기 위해선 숙소의 유형, 편의 시설 제공 여부 등 많은 단계를 거쳐야하므로 불편하다. 본 논문은 호스트가 보다 편리하게 자신의 숙소에 최적화된 가격을 알 수 있도록 하는 '숙소의 적정 가격 결정을 위한 웹 사이트'를 개발하였다. 이 웹을 통해 호스트는 더 간편하게 자신의 숙소에 대한 적정 가격을 알고 가격 산정 시 참고할 수 있다.

키워드: 숙소 가격 추정 (Estimation of Accommodation Price),
K-최근접 이웃 알고리즘 (K-nearest neighbor algorithm)

I. 서론

사람들이 여행에 대한 관심도가 높아지면서[1] 숙박 플랫폼을 통해 숙소를 예약하기도 하고, 자신이 직접 호스트가 되어 숙소를 제공하기도 한다. 숙소 요금 산정은 전적으로 호스트의 권한이지만 숙박 플랫폼이 숙소에 맞는 최적화된 가격을 제공하여 산정에 도움을 준다. 하지만 호스트가 이러한 최적화된 가격을 알아보고자 할 때 등록 절차인 숙소의 유형, 편의시설 제공 여부 등 많은 단계를 거친 후에 최적화된 가격을 알 수 있어 불편하다. 이러한 점을 개선하기 위해서 다양한 숙소의 데이터를 수집하여 모델을 만들고 호스트가 숙소에 대한 정보가 포함된 CSV 파일을 업로드하면 기존의 모델을 바탕으로 숙소 가격을 추정하는 웹 사이트를 개발하고자 한다.

II. 숙소 가격 추정

가격 추정 모델을 만들기 위해 'Inside Airbnb'[2] 데이터를 사용한다. 'Inside Airbnb'는 세계 주요 대도시 숙소의 고유 아이디ID, 침실 수, 화장실 수, 가격 등 다양한 데이터를 제공하고 있다. 본 논문은 2020년 02월 미국의 일부 21개 지역의 데이터를 사용하였다. 그리고 데이터가 분석에 적합하도록 전처리 과정을 거친다.

id	host_id	accommodates	bathrooms	bedrooms	beds	price
38585	165529	2	1	1	2	$50.00
80905	427027	12	2	3	6	$765.00

Fig. 1. 'InsideAirbnb' 제공 데이터[2]

id	host_id	accommodates	bathrooms	bedrooms	beds	price
38585.0	165529.0	2.0	1.0	1.0	2.0	50.0
80905.0	427027.0	12.0	2.0	3.0	6.0	765.0

Fig. 2. 전처리 후 데이터

그림1에서 그림 2처럼 전처리 과정을 통해 price 열에 '$'를 없애고, 평균을 계산할 때 정교한 값을 얻기 위하여 모든 데이터의 형식을 실수형으로 형 변환한다.

데이터 분석은 K-nearest neighbor - Regression (이하. KNN)[3] 알고리즘을 사용한다. K-nearest neighbor - Regression이란 K-최근접 이웃 알고리즘의 회귀방식으로, 구하고자 하는 데이터와 가장 유사한 데이터 k개를 선택하고 그 값의 평균을 계산하는 알고리즘이다. 이러한 알고리즘을 이용한 모델을 평가하기 위해 Root Mean Square Error (이하 RMSE)[4]를 사용한다. RMSE란 평균 제곱근 편차로

모델이 예측한 값과 실제 값의 차이를 다룰 때 흔히 사용하는 측도로, 실제 값에서 예측 값의 차이를 기반으로 하기 때문에 RMSE 값이 작을수록 모델의 성능이 좋다.

학습과 테스트는 12,494개(전체 데이터[2]/8)를 9:1 비율로, KNN의 k는 5로 진행한다. 전체 데이터[2]에 1/8만 사용하고, k를 5로 주는 이유는 RMSE 값이 가장 낮은 값이 나오기 때문이다. 다음은 임의의 수로 데이터를 나눈 값을 이용하여 가격을 추정하였을 때 RMSE를 나타낸 그래프이다.

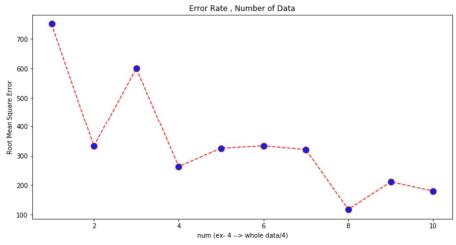

Fig. 4. matplotlib[5]으로 표현한 k가 default값(5)이고 임의의 수(x)로 데이터를 나눈 값을 통한 RMSE

그림 4를 통해 k=5, 전체 데이터를 8로 나누었을 때, RMSE 값이 가장 작게 계산됨을 알 수 있다.

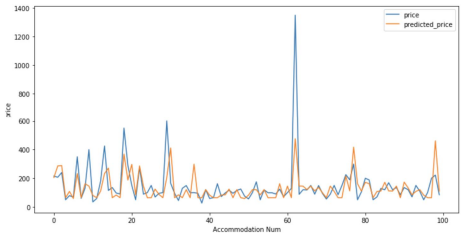

Fig. 5. matplotlib[5]으로 표현한 실제 가격과 추정 가격을 통한 가격의 차이

그림 5을 통해 해당 모델을 사용하였을 때 실제 가격과 비슷한 추정 가격이 나옴을 알 수 있다.

III. 숙소의 적정 가격 결정을 위한 웹 사이트

웹을 구성할 때는 Spring Framework로 웹 개발을 진행한다. 호스트가 보게 되는 웹 페이지 화면은 파일 업로드 페이지와 결과를 보여주는 페이지로 총 2가지이다. 첫 번째 파일 업로드 페이지의 주요 기능은 파일 업로드 기능이다. 호스트가 [파일 선택]이라는 버튼을 클릭하여 파일을 선택하고 [업로드]를 클릭하면 Controller에서 Command를 실행할 수 있는 Java Runtime (exec)를 통해 Python 파일을 실행시킨다. 이때 Python 파일에선 데이터와 숙소 가격 추정 모델을 이용하여 숙소의 가격을 추정하고 결과를 DB에 저장하는 기능을 수행한다. Python에서의 실행을 마치면 다시 Controller로 돌아와 DB에 저장된 결과를 Spring Framework로 구현한 웹페이지를 통해 차트로 호스트의 숙소에 대한 적정 가격을 나타낸다.

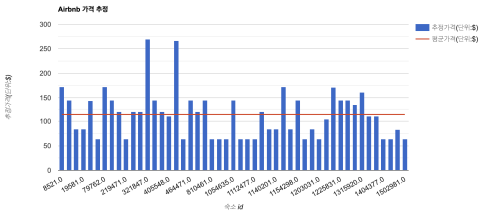

Fig. 6. GoogleChart[6]를 이용한 추정가격 차트

그림 9 차트의 x는 숙소 ID를 y는 그에 따른 각각의 결과 값은 막대 형태로, 전체의 평균값은 선을 이용하여 표현한다.

IV. 결론

본 연구를 통해 호스트는 복잡한 과정 없이 CSV파일을 업로드하는 것만으로 호스트의 적정 숙소 가격을 추정할 수 있다. 이후에는 날씨, 비수기, 성수기와 같은 숙소에 대한 데이터를 추가함으로써 더욱 정교한 숙소 가격을 추정할 수 있을 것으로 기대된다.

REFERENCES

[1] Statistical data for Korea Tourism http://kto.visitkorea.or.kr/kor/ktom/stat/stat.kto
[2] Inside Airbnb, http://insideairbnb.com/get-the-data.html
[3] K-nearest neighbor (KNN) https://www.javatpoint.com/k-nearest-neighbor-algorithm-for-machine-learning
[4] Root Mean Squared Error(RMSE) http://statweb.stanford.edu/~susan/courses/s60/split/node60.html
[5] Matplotlib plot, https://matplotlib.org/
[6] Google Charts, https://developers.google.com/chart