

효율적인 중고거래 메타서비스

황세웅*, 임민택*, 홍현기^o, 황훈태*, 박성현*, 최영규*, 황석형*, 김수환*

*선문대학교 AI소프트웨어학과,

^o선문대학교 AI소프트웨어학과

e-mail: {sewoong*, lmtna99*, gnsxo2484*, asdpark12*, mpd253123*, shwang*, kimssoohwan*}@sunmoon.ac.kr, hk7286@sunmoon.ac.kr^o,

An Efficient Second-hand transaction meta-services

Sewoong Hwang*, Min-Taek LIm*, Hyun-Ki Hong^o, Hun-Tae Hwang*,

Sung-Hyun Park*, Young-Kyu Choi*, Suk-Hyung Hwang*, Soo-Hwan Kim*

*Department of Artificial Intelligence and Software Technology,

^oDepartment of Artificial Intelligence and Software Technology

● 요약 ●

본 논문에서는 기존 중고거래 플랫폼들의 불편한 점들을 해소하고 사용자들이 효율적이고 편리한 중고거래를 할 수 있도록 도와주는 플랫폼을 개발했다. 조사를 통해 기존 중고거래 플랫폼은 허위 매물, 시세 파악의 어려움, 사기 피해 등의 문제점이 존재한다는 사실을 인식했다. 문제 해결을 위해 파이썬을 활용하여 주요 중고거래 플랫폼의 상품 데이터를 수집했다. 이에 IQR을 적용하여 가격의 이상치를 판별했다. 가격 비교와 허위 매물 판별이 용이하게 되는 장점이 있다. 또한 이상치를 제거한 상품들의 시세를 계산하여 데이터를 차트로 시각화했다. 플랫폼과 지역마다 상이한 중고 상품의 신뢰성 있는 시세를 파악할 수 있고 중고거래 사기 피해를 방지할 수 있도록 사용자에게 주요 사기 수법, 뉴스 등의 정보를 제공한다.

키워드: 중고거래(Second-hand transaction), 중고상품시세(Second-hand product market price), IQR(InterQuartile Range)

I. Introduction

중고거래 시장은 빠른 속도로 성장하고 있다. 업계 추산 2008년 약4조 원이었던 국내 중고 소비시장 규모는 2020년 20조 원으로 증가했다.[1] 중고거래 앱 이용자는 19년도 기준 337만 명을 돌파했다. 또한, 국내 중고 거래 앱 월간 사용자는 1,700만 명을 넘어섰다.[2] 성장하고 있는 시장이지만 해결해야 하는 문제들이 있었다. 주요 중고거래 플랫폼을 분석한 논문[3]에 따르면 사용자는 다음과 같은 불편함을 느꼈다고 한다. 첫째, 허위 매물이 많아 중고 상품의 선택이 힘들다. 둘째, 중고 상품의 정확한 시세를 알기 어렵다. 셋째, 사기 수법이 다양하여 거래하기 꺼려진다. 중고 거래 시장이 성장하는 만큼, 서비스 이용의 불편함 또한 증가한다. 따라서 기존 중고 거래 플랫폼의 문제점을 해소하고 사용자가 더 효율적이고 안전하게 중고 거래를 할 수 있도록 허위 매물 탐지 연구를 진행하였다.

본 논문은 가격 데이터와 IQR을 활용하여 가격의 이상치를 판별하는 방식을 제안한다. 해당 방법론은 상품가격 데이터에 IQR을 적용하여 해당 상품가격 데이터가 Q1과 Q3 범위 안에 속하지 않으면 해당 상품을 이상치, 즉 허위 매물로 판단하는 방식을 적용하였다.

이를 이용해 허위 매물을 제거한 상품들의 시세를 계산하여 시각화 자료로 사용자에게 제공한다. 따라서 본 논문은 해당 플랫폼의 구현과 정에서 생긴 요구조건과 그에 대한 방법론을 구체적으로 제시한다.

II. Preliminaries

1. Related works

1.1 기존 중고거래 플랫폼 문제 현황

한국의 대표적인 3사 중고 플랫폼 이용자들은 플랫폼별 다양한 페인 포인트를 가지고 있다[3]. 중고나라 앱의 경우 불필요한 광고, 직거래에 대한 불편함, 그리고 허위 매물을 페인 포인트로 도출했고 번개장터의 경우 거래 수수료, 매물 종류와 수 그리고 검색과 문의 단계에서 페인 포인트로 도출할 수 있었다. 마지막으로 당근마켓의 경우 다른 지역 매물을 확인할 수 없다는 어려움과 다른 플랫폼과 마찬가지로 허위 매물에 대한 불편함이 페인 포인트로 도출되었다.

본 필자는 3사 플랫폼의 공통적인 불편함인 허위 매물의 불편함을 해소하고자 연구를 진행하였다.

1.2 단변량 자료에서 이상치 탐색의 다양한 방법

단변량 자료에서 이상치 탐색 방법은 변수가 하나인 자료에서 이상치를 탐색하는 방법이다. 단변량 자료에서 이상치 탐색 방법으로는 표준화 점수(Z-score), 통계적 가설 검정, 사분위수 범위 등이 있다. 표준화 점수를 활용한 이상치 탐색 방법에서, 표준화 점수는 평균이 이고, 표준편차가 인 정규분포를 따르는 관측치들이 자료의 중심(평균)에서 얼마나 떨어져 있는지를 반영한다. 통계적 가설검정을 활용한 이상치 탐색은 최솟값 혹은 최댓값의 이상치 여부에 대한 검정으로, 이상치로 판단된 관측치를 제외해 나가면서 더 이상 이상치가 존재하지 않을 때까지 반복적으로 검정을 수행하여 이상치를 정의하는 방법이다[4]. 통계적 가설검정을 활용한 이상치 탐색 방법은 디스 Q검정(Dixon Q-test), Grubbs test, Generalized ESD(extreme studentized deviate) 검정, 카이제곱 검정(Chi-square test) 등이 있다. 사분위수 범위(interquartile range, IQR)를 활용한 이상치 탐색은 상자그림에서 사분위수 범위의 1.5배를 초과하는 관측치는 이상치로 정의하는 방법이다.

1.3 IQR(Interquartile range)

통계량에서 사분위간 범위(IQR)는 데이터의 산포를 나타내는 통계적 산포의 척도이다. IQR은 Fig 1과 같이 75번째 백분위수와 25번째 백분위수의 차이로 정의된다. IQR을 계산하기 위해, 데이터를 4개의 분위로 나누는데 이러한 사분위수는 Q1(제 1사분위 수), Q2(제 2사분위 수), Q3(제 3사분위 수)로 표시한다. IQR은 제 3사분위 수와 제 1사분위 수 사이의 값으로 정의함으로 $IQR=Q3 - Q1$ 이다[5].

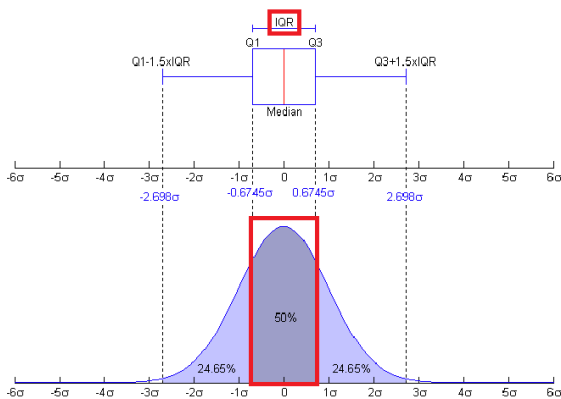


Fig. 1. IQR

III. The Proposed Scheme

본 연구에서 개발한 시스템의 아키텍처는 Fig 2와 같다.



Fig. 2. ResellViewer Architecture

1. 데이터수집 및 개요

중고나라, 반개장터, 당근마켓 3사 플랫폼에서 게시물 명, 게시물 주소, 가격, 위치, 이미지 주소, 업로드 시간 데이터를 실시간 크롤링 기법으로 수집하였다. 이렇게 수집된 중고 상품 데이터는 데이터베이스에 저장하여 데이터 분석 및 가공 작업에 이용한다.

2. 데이터 분석 및 가공

수집한 데이터를 IQR기법을 사용하여 이상치 판별 작업을 진행하였다. IQR(Interquartile range)기법은 제 3분위 수에서 1분위 수 사이의 값을 이용하여 이상치를 판별하는 기법이다. 여러 번의 테스트 과정을 거쳐 “시세 이하 범위 < $Q1 - 0.4 * IQR \leq$ 시세 범위 $\leq Q3 + 0.4 * IQR <$ 시세 이상 범위” 수식을 적용하였을 때 시세를 가장 잘 판별하는 것을 확인했다.

3. 판별한 이상치를 시각화 및 제거

평균보다 많이 떨어져 있는 가격들을 이상치(평균 이하, 평균 이상)로 판별하여마켓에서 좌측 상단 라벨을 통해 시각화 자료 Fig 4를 제공한다. 프론트 엔드는 백엔드 서버로부터 IQR기법이 적용된 가격의 이상치 여부인 ‘outlier’를 JSON 형식의 데이터를 받아 시각화를 진행한다. JSON 형식은 Fig 3과 같다.

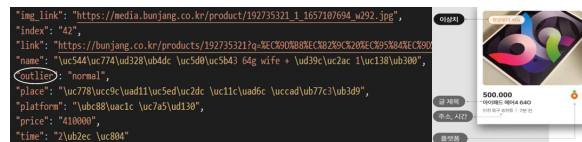


Fig. 3. Outlier

가격의 이상치를 제거한 범위의 중앙값을 시세로 계산하여 차트로 시각화 자료 Fig 5를 제공한다. 매물이 없는 날짜의 경우에는 전날 시세 평균으로 대체하여 결측치를 처리했다.

4. 사용자 화면

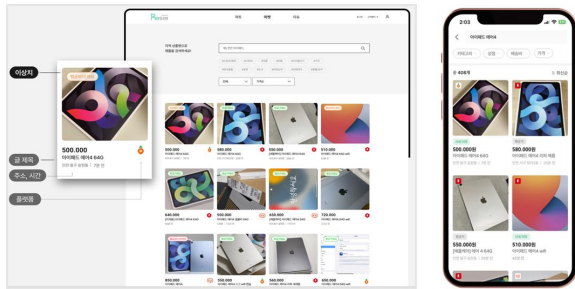


Fig. 4. ResellViewer Market Page

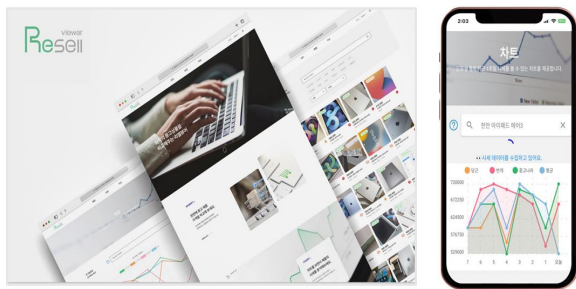


Fig. 5. ResellViewer Web and Chart

IV. Conclusions

본 논문은 기존 중고 거래 플랫폼의 불편함을 해소하고 사용자가 효율적이고 편리하게, 또한 안전하게 중고 거래를 할 수 있는 플랫폼을 만들기 위한 목적으로 쓰였다. 본 연구에서 필자는 중고거래 이용자들이에게 합리적인 가격선의 상품 제안과 동시에 평균에서 크게 떨어져 있는 가격선의 상품들을 한눈에 시각화하기 위해 IQR을 활용하여 적용하였다. 이에 따른 결과로 IQR을 통한 이상치로 판별된 매물에 대해 약95%의 이상치 판별률을 보였다. 이상치 판별 정확도를 확인해 보기 위해 “아이패드 에어4” 키워드로 성능평가를 진행하였다.

Table 1. “아이패드 에어4” 이상치 판별 성능평가

	개수	일치	불일치	정확도
High	2	2	0	100%
Normal	47	47	0	100%
Low	13	11	2	84.6%
total	61	59	2	96.7%

현재 중고상품의 시세를 세 단계로 분류한 상태로 사용자에게 제공하고 있다. 필자는 가격 선 분류를 세 단계로 분류하는 것만으로는 이상치 판별에 한계점이 있다고 판단하여 추후 개선 사항으로 다섯 단계 분류를 계획하고 Machine Learning 알고리즘을 적용할 예정이다.

REFERENCES

[1] “Perhaps... Are you a carrot?” Growth of second-hand transaction platform Carrot Market, Nielsen[Website]. (2022, Apr 12). http://www.koreanclick.com/insights/new_sletter_view.html?code=buzzword&id=612&page=1&utm_source=board&utm_medium=board&utm_campaign=buzzword&utm_content=20210322

[2] Android Mobile Second-hand transaction application usage status, Nielsen[Website]. (2022, Apr 12). http://www.koreanclick.com/insights/newsletter_view.html?code=digitalnow&id=521&page=1&utm_source=board&utm_medium=board&utm_campaign=digitalnow&utm_content=20190412

[3] Sung Phil Kim. (2021). A Study on the User Structure Analysis of Second Hand Market Apps -Focusing on Joongonara, Bungaejangter, Danggeunmarket-. Journal of the Korea Academia-Industrial cooperation Society, 22(7), 449-458.

[4] J.H. Kim. "Statistical methods for outlier detection". HIRA Policy Brief, vol. 14, pp.49-56, 2020.

[5] Zhang, Y., Kim, N. H., & Haftka, R. T. (2020). General-surrogate adaptive sampling using interquartile range for design space exploration. Journal of Mechanical Design, 142(5).