**영화 리뷰에서 사용자의 전문성을 고려한 경험기반 신뢰도**

박강민

서울과학기술대학교 데이터사이언스학과

**Historical Credibility for Movie Reviews considering User’s Expertise**

Gang-Min Park

Data Science, Seoul National University of Science and Technology

**요약**

온라인 상의 영화 리뷰는 영화를 관람한 사람이 작성하는 주관적인 평가 또는 감상을 의미하며 관객이 영화를 선택하는데 중요한 지표 역할을 한다. 어떤 영화 리뷰어들은 전문적으로 수만건의 영화 리뷰를 작서아긱도 하지만 일부 영화 리뷰어들은 극단적이고, 악의적인 리뷰를 작성하여 영화 리뷰를 지표로 사용하는데 혼동을 줄 수 있다. 과거 영화 데이터를 기반으로 영화 리뷰의 신뢰성을 판단하는 historical credibility는 리뷰 데이터의 rating, sentiment score 및 correlation score를 활용하여 리뷰의 신뢰성을 판단한다. 하지만 각각의 historical credibility는 영화 리뷰어의 영화에 대한 기호나 시청 경험, 리뷰작성 경험등을 전혀 고려하지 않았기 때문에 정확한 리뷰 신뢰도 판단을 위한 근거로 사용될 수 없다. 따라서 본 논문에서는 영화 리뷰어의 전문성을 고려한 historical credibility를 제안한다. 영화 리뷰어의 전문성은 리뷰어가 과거 리뷰한 영화의 장르를 고려한 장르 전문성과 리뷰어의 리뷰 수를 고려한 리뷰 전문성으로 구성된다. 실험을 통해 historical credibility를 기준으로 trusted review를 추출하고 전문성 지표를 반영 및 rating을 재평가하여 리뷰어의 전문성이 반영된 rating지표를 얻을 수 있음을 확인하였다.

1. **서론**

영화 리뷰는 영화를 관람한 사람이 오프라인이나 온라인 상에서 작성하는 주관적인 평가 또는 감상을 의미하며 리뷰어의 성향, 기호, 과거 시청 영화 또는 리뷰작성 경험 등에 영향을 받아 작성된다. 온라인 영화 리뷰의 경우 일반적으로 두가지 유형의 데이터로 구성된다: 1) textual reviews and 2) ratings. Textual review는 영화에 대한 사용자의 감상을 나타내는 질적 평가이며 rating은 영화에 대해 수치화 된 평점이다. 그림 1은 네이버에서 제공하는 영화 리뷰 페이지로, 140자 제한의 textual review와 0에서 10점 사이의 평점을 제공한다. 영화에 대한 사용자의 textual review와 rating은 상호 독립적이지 않고 밀접한 관계를 가진다. 예를 들어, 영화에 대해 호의적인 내용의 textual review일수록 rating또한 높은 점수대를 가진다. 따라서, 영화에 대한 가장 간단한 평가를 ratings으로, 그 다음을 textual review로 정의할 수 있다.

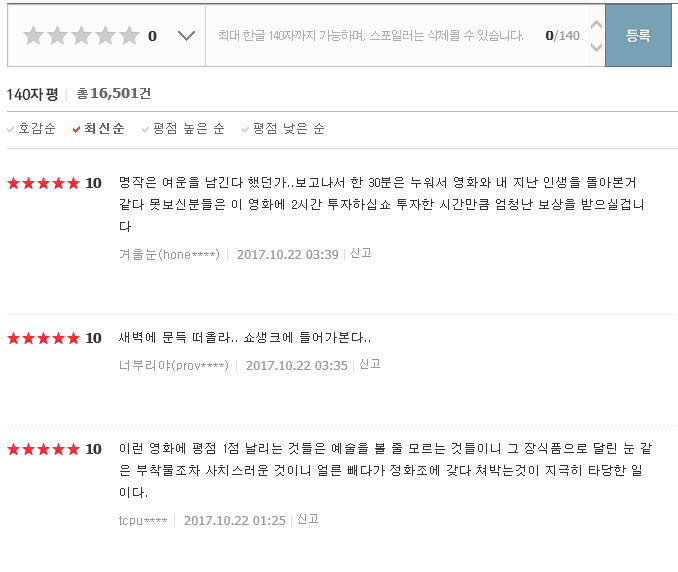


그림 1. 네이버 영화 리뷰 페이지

Textual review와 rating 기반의 영화 리뷰는 관객이 영화를 선택하는데 중요한 지표 역할을 한다. 어떤 영화 리뷰어들은 전문적으로 수만건의 영화 리뷰를 작성하기도 하지만 일부 영화 리뷰어들은 극단적이고, 악의적인 리뷰를 작성하여 영화 리뷰를 지표로 사용하는데 혼동을 줄 수 있다. 이 경우, 영화 내용과 무관한 textual review와 rating일 수 있으며 둘사이의 관계 또한 독립적일 수 있다. 따라서 영화 리뷰를 유효한 지표로 사용하기 위해서는 영화 리뷰의 신뢰성을 평가하는 것이 필요하다. 영화 리뷰의 신뢰성은 전문가에 의해서도 판단되기 어려운 문제이며, 많은 영화 리뷰 수로 인해 수작업으로 할 경우 많은 시간과 노력을 요구한다. Kim et al. [1]는 과거 리뷰데이터로부터 영화 리뷰의 신뢰도를 판단하기 위한 3가지 종류의 historical credibility feature를 제안하였다: rating, sentiment scores, correlation scores. 각각의 historical credibility features들은 rating, textual review의 감성 점수, rating과 textual review 감성점수의 상관관계의 표준편차를 수치화 하며, 편차가 클수록 신뢰할 수 있는 사용자로 판단한다. 하지만 historical credibility feature는 영화 리뷰어의 영화에 대한 기호나 시청 경험, 리뷰작성 경험 등을 전혀 고려하지 않았기 때문에 정확한 리뷰 신뢰도 판단을 위한 근거로 사용될 수 없다. 예를 들어, 특정 장르의 영화만 보는 리뷰어의 다른 장르 영화에 대한 리뷰는 편향된 결과일 수 있으며 리뷰를 수십건 작성한 리뷰어와 수천건 작성한 리뷰어 사이에는 명확한 신뢰도 차이가 존재할 수 있다.

본 논문에서는 영화 리뷰어의 전문성을 고려한 historical credibility를 제안한다. 영화 리뷰어의 전문성은 과거 리뷰한 영화의 장르 편향도를 고려한 장르 전문성 및 리뷰자체의 수를 고려한 리뷰 전문성으로 구성된다. 사용자의 전문성을 고려한 historical credibility는 기존의 리뷰 신뢰도 평가를 보완할 뿐 아니라 장르 및 리뷰 전문성을 고려한 rating 재산정을 통해 신뢰할 수 있는 rating을 얻을 수 있다.

본 논문의 구성은 다음과 같다. 2장에서는 선행연구에 대해 설명한다. 3장에서는 제안하는 방법에 대해 설명한다. 4장에서는 실험 결과를 설명한다. 5장에서는 본 논문의 결론과 향후 연구를 설명한다.

1. **선행연구**

Kim et al. [1]는 과거 영화 리뷰 데이터를 기반으로 영화 리뷰의 신뢰성을 판단하는 historical credibility를 제안하였다. 영화 리뷰의 신뢰성 판단을 위한 기존 연구에서 주로 사용한 신뢰성 지표, helpfulness voting은 영화 리뷰에 다른 사용자들이 남긴 ‘좋아요’, ‘싫어요’ 투표 수를 기반으로 신뢰성을 평가한다. 하지만 helpfulness voting은 투표를 받기 위한 일정 시간이 요구되고, 사용자의 리뷰가 노출된 시기에 따라 다른 평가를 받을 수 있으며 영화의 리뷰어 수에 따라 수치가 일정하지 않는 등 신뢰성을 평가하기 위한 지표로는 부적합하다. 특히 그림 2에서 기존연구에서 helpfulness voting 기반 상/하위 리뷰들을 신뢰 여부로 결정하고 감정분석을 수행했을 때 실제 e-commerce의 거짓 품목에 대한 가짜 리뷰의 분포를 반영하지 못했다[2].

텍스트, 라인, 그래프, 도표이(가) 표시된 사진

자동 생성된 설명

(a) Fraud and normal items’ comments on e-commerce [2]. (b) Trusted and distrusted movie reviews classified by helpfulness votes.

그림 2. E-commerce 리뷰와 영화 리뷰의 감정분석 결과 비교

따라서 Kim et al. [1]는 효과적인 리뷰의 신뢰성 판단 및 분류를 위해 3가지 종류의 historical credibility 제안하였다: 1) rating, 2) sentiment score and 3) correlation. 1) rating은 리뷰어가 작성한 과거 리뷰 데이터의 rating 표준편차를 기반으로 리뷰의 신뢰도를 판단한다. 2) sentiment score는 textual review에서 감성점수를 추출하여 감성점수의 표준편차를 기반으로 리뷰의 신뢰도를 판단한다. 3) correlation은 rating 및 sentiment score가 일치하는지 여부를 판단하기 위해 correlation score를 계산하여 수치가 클수록 신뢰할 수 있는 사용자로 판단한다. 이때 rating과 sentiment score의 경우, 과거 시청한 영화들에 대해 비슷한 평가를 한 리뷰어 보다 다양한 평가를 한 리뷰어가 더욱 신뢰할 수 있다는 가정으로, 표준편차가 큰 리뷰어를 신뢰할 수 있는 리뷰어로 판단한다. Kim et al. [1]이 제안하는 historical credibility의 추출 과정은 그림 3과 같다.

텍스트, 폰트, 번호, 스크린샷이(가) 표시된 사진

자동 생성된 설명

그림 3. Historical credibility 추출 프레임워크

Kim et al. [1]이 제안하는 historical credibility는 Weng et al. [2]의 실제 거짓품목에 대한 리뷰 데이터의 분포를 따라가는데 목적이 있다. 따라서 validation을 위한 distribution analysis, statistical analysis 및 correlation analysis가 수행되었으며, 각각의 분석에서 historical credibility를 기반으로 한 사용자의 분류 및 분석 결과가 helpfulness voting을 기반으로 한 결과에 비해 실제 데이터의 분포를 더욱 잘 나타내었다. 분석결과 distrusted 리뷰는 대체로 긍정적이고 영화의 종류와 관계없이 구별하기 어려우며 일관성이 없고 rating과 sentiment score사이의 상관관계가 적은 것으로 나타났다. 이를 통해 Kim et al. [1]은 historical credibility가 리뷰의 신뢰도를 판단하는데 유효하다는 것을 증명하였다.

1. **제안하는 방법**

본 연구에서는 사용자의 과거 영화리뷰 데이터에서 영화의 장르 편향도를 고려한 장르 전문성 및 리뷰 자체의 수를 고려한 리뷰 전문성을 추정하여 historical credibility를 추정한다.

* 1. **장르 전문성**

특정 사용자의 장르 편향도를 고려하기 위해서는 사용자의 과거 영화리뷰 데이터와 영화의 장르를 매칭시킬 필요가 있다. 본 연구에서는 사용자의 영화 리뷰 데이터와 영화 장르 데이터를 매칭시키기 위해 KOBIS 영화 정보 데이터[[1]](#footnote-1)에서 장르 데이터를 수집하여 사용하였다. 수집된 영화 리뷰데이터에서 리뷰된 37786개의 영화 중 장르 카테고리가 존재하지 않는 16016개를 제외한 21770개의 영화만이 분석에 사용되었다. 매칭된 특정 사용자의 영화 리뷰 데이터와 장르 데이터는 사용자의 특정 장르에 대한 선호도를 반영하기 위해 전체 데이터에 대해 min-max 정규화를 수행한다. 그림 4는 특정 사용자의 과거 영화리뷰 데이터에서 해당하는 영화의 장르를 매칭한 결과이다. 장르가 두개 이상인 영화의 경우 두 장르 모두 점수를 부여한다. 그림 4의 좌측은 총 21개의 영화 장르로 리뷰데이터를 매칭한 결과이며 그림 4 우측의 결과처럼 전체 장르에 대해 정규화 하여 장르 전문성 feature로 활용한다. 특정 사용자의 리뷰데이터에 장르 카테고리가 존재하지 않는 영화가 있을 경우 ‘NAN’ 항목에 해당하며 정규화 시 제외한다.

* 1. **리뷰 수 전문성**

특정 사용자의 리뷰 전문성을 고려하기 위해서는 사용자의 과거 영화리뷰 데이터의 수를 반영하여 feature로 활용한다. 장르 전문성 feature만 존재할 경우 특정 사용자의 장르 선호도에서 우연성을 배제할 수 없기 때문에 리뷰 전문성 feature를 통해 장르 전문성 feature를 보충하는 효과가 있다. 또한 리뷰 전문성 feature는 사용자가 얼마나 신뢰할 수 있는 리뷰를 작성하는지에 대한 지표로 해석할 수 있다. 그림 4 우측의 review\_num은 특정 사용자의 과거 영화리뷰 데이터에서 ‘NAN’항목을 제외한 총 리뷰 수를 의미한다.

텍스트, 영수증, 스크린샷이(가) 표시된 사진

자동 생성된 설명

그림 4. 특정 사용자의 영화 장르 맵핑 및 정규화

**3.3 feature 생성**

본 논문에서 제안하는 영화 리뷰데이터의 features는 3가지 historical credibility features인 rating, sentiment score, correlations core에 기존의 textual review, 장르 전문성 category, 리뷰 전문성 reviewNum으로 구성되어 있다. 그림 5는 특정 영화에 대한 리뷰 feature들을 나타낸다. 실험을 위해 수집한 리뷰데이터에 존재하는 사용자의 수는 약 22만명 이며 사용자별 최대 1만건의 리뷰가 수집되었다. 그림 6는 사용자 별 리뷰 수를 나타내며 분석에서는 3개 이상의 리뷰를 작성한 사용자의 리뷰데이터만을 사용하였다. 리뷰 전문성의 경우 outlier가 극심하기 때문에 특정 영화에 대한 reviewNum을 중앙값과 IQR을 활용한 robust scale을 사용하여 정규화를 수행하였다.

텍스트, 영수증, 스크린샷, 폰트이(가) 표시된 사진

자동 생성된 설명

그림 5. 특정 영화에 대한 리뷰 feature

라인, 도표, 그래프, 텍스트이(가) 표시된 사진

자동 생성된 설명

그림 6. 사용자 별 리뷰 수

1. **실험**
   1. **데이터**

본 논문에서는 제안하는 사용자의 전문성을 고려한 historical credibility를 위해 특정 영화에 대한 리뷰의 rating 및 textual review를 수집하였다. 표 1은 리뷰데이터 수집대상으로 선정된 7개의 영화와 각 영화에 해당하는 장르를 나타낸다. 장르는 총 21개가 존재한다. 사용자의 과거 리뷰 데이터에 존재하는 영화는 총 37786개이며 이 중 장르 항목이 없는 16016개를 제외한 21770개의 영화에 해당하는 리뷰 데이터만을 사용하였다.

표 1. 영화 및 장르 매칭

텍스트, 폰트, 스크린샷, 번호이(가) 표시된 사진

자동 생성된 설명

실험을 위해 Kim et al. [1]의 제안에 따라 각 영화 별 3가 historical credibility(rating, sentiment score, correlation score)를 기준으로 상/하위 0.3에 해당하는 리뷰 데이터를 trusted review/ distrusted review로 분류한다.

표 2. 사용자의 전문성 지표를 고려한 rating 재평가

텍스트, 스크린샷, 번호, 폰트이(가) 표시된 사진

자동 생성된 설명

* 1. **rating 재평가**

영화 리뷰 사용자의 장르 전문성 및 리뷰 전문성을 고려한다면 영화에 대해 좀더 정확한 rating을 얻을 수 있다. 표 2는 사용자의 전문성 지표를 고려하여 재평가된 rating을 historical credibility 기반 distrusted 리뷰 및 trusted 리뷰의 rating과 비교하였다. 7개의 영화에 대해 historical credibility는 rating, sentiment 및 correlation 기준 trusted/distrusted 리뷰에 대한 rating을 평가하였으며 전문성 반영 rating의 경우 장르 전문성과 리뷰 수 전문성을 sum scale하여 rating과 가중평균을 계산한다. 장르 전문성의 경우 리뷰어가 시청한 영화 장르에 대한 이해도를 반영하기 위해 수집된 영화의 장르와 일치하는 장르 전문성 수치를 곱하여 rating을 계산한다. 리뷰 수 전문성의 경우 장르 전문성에 비해 outlier의 영향이 크기 때문에 전문적인 리뷰어의 영향을 크게 반영할 수 있다. 따라서 historical credibility 기준 trusted로 분류된 리뷰들의 rating이 distrusted에 비해 좀 더 정확한 rating이라고 해석할 수 있으며 전문성이 반영된 rating이 더욱 정확한 rating이라고 해석할 수 있다.

**5. 결론**

본 논문에서는 영화 리뷰어의 전문성을 고려한 historical credibility를 제안한다. 영화 리뷰어의 전문성은 리뷰어가 과거 리뷰한 영화의 장르를 고려한 장르 전문성과 리뷰어의 리뷰 수를 고려한 리뷰 전문성으로 구성된다. 전문성 지표는 기존의 historical credibility와 함께 신뢰할 수 있는 리뷰를 구별해 내는데 도움이 될 수 있다. 실험에서 historical credibility를 기준으로 trusted review를 추출하고 전문성 지표를 반영 및 rating을 재평가하여 리뷰어의 전문성이 반영된 rating지표를 얻을 수 있음을 확인하였다.

본 연구에서는 제안하는 전문성 지표의 효과를 확인하기 위해 historical credibility 기반 trusted review의 rating을 재평가하였지만, 향후 전문성 지표를 trusted review를 구별하는데 활용하여 더욱 정확한 리뷰 신뢰성 판단을 제안하고자 한다.

**6. 참고문헌**

[1] Kim, M. S., Lim, B. Y., Shin, H. S., & Kwon, H. Y. (2023). Historical credibility for movie reviews and its application to weakly supervised classification. *Information Sciences*, *630*, 325-340.

[2] Weng, H., Ji, S., Duan, F., Li, Z., Chen, J., He, Q., & Wang, T. (2019, April). Cats: cross-platform e-commerce fraud detection. In *2019 ieee 35th international conference on data engineering (icde)* (pp. 1874-1885). IEEE.

1. https://www.kobis.or.kr/kobis/business/mast/mvie/searchMovieList.do [↑](#footnote-ref-1)