

품사 정보를 활용한 크라우드 펀딩 성공 여부 예측

안유나*, 최지우*, 김웅기*, 김서영*

Predicting Crowd Funding Success Using Part of speech Information

Ahn Yuna*, and Choi Jiwoo*, and Kim Woongki*, and Kim Seo Young*

요 약

크라우드 펀딩 플랫폼의 등장으로 인해 투자영역의 폭이 넓어지고 있다. 이에 따라 투자자의 투자 실패를 예방하고자 펀딩 프로젝트의 성공 여부 예측에 관한 연구가 꾸준히 진행되고 있다. 본 연구에서는 크라우드 펀딩 성공 여부 예측에 품사 정보가 어떠한 영향을 미치는지에 연구해 보고자 한다. 우선, 크라우드 펀딩 플랫폼에서 제공하는 정보인 펀딩 목표 금액, 펀딩 기간, 이미지 개수, 제목 길이와 메이커의 이전 프로젝트 성공 및 실패 개수와 신규 메이커 여부로 모델 구축하였다. 해당 모델에 파이썬 KoNLPy 라이브러리로 품사를 태깅하여 얻은 품사 출현 빈도 상위 5개의 품사를 변수로 추가하여 모델을 구축한 후 두 모델을 비교하여 품사 정보가 가지는 영향을 파악해보았다. 3가지 분류 기법 LightGBM, GradientBoosting, AdaBoost을 사용하여 펀딩 정보만을 고려한 모델과 펀딩 정보에 품사 정보를 더한 모델을 비교하였다. 이를 통하여 품사 정보가 크라우드 펀딩 성공 여부 예측에 근소하지만, 긍정적인 영향을 미친다는 사실을 확인할 수 있었다.

Abstract

With the advent of crowdfunding platforms, the scope of investment is widening, and research on the prediction of success of funding projects is steadily progressing to prevent investors from failing to invest. The purpose of this study is to investigate the effect of part-of-speech information on the prediction of crowd funding success. First of all, we built a model based on the funding target amount, the pinning period, the number of images, the title length, the number of successes and failures of the previous project of the maker, and whether it is a new maker. The model was constructed by adding the top five parts of speech frequency obtained by tagging the parts of speech with the Python KoNLPy library as variables, and then the effects of the part of speech information were compared with the two models. Three classification methods, LightGBM, GradientBoosting, and AdaBoost, were used to compare the model considering only funding information with the model adding part of speech information to the funding information. Through this, it was confirmed that the part-of-speech information has a small but positive effect on the prediction of crowdfunding success.

Key words

crowd funding, machine learning, Wadiz, prediction success, classifier, text mining

I. 서 론

크라우드 펀딩(Crowd Funding)이란 Crowd(대중), Funding(자금조달)의 합성어로 여러 경로를 통해서 다수의 투자자로부터 투자금을 모으는 것을 뜻한다. 이를 통해 투자금을 유치하기 위해, 높고 어려웠던 사업 제안 절차와 과정이 단순화되었으며, 소규모 투자집단들의 투자영역의 폭을 넓힐 수 있게 되었다. [1] [2] 투자의 폭이 넓어짐에 따라 크라우드 펀딩 프로젝트에 대한 투자자의 투자 실패로 인한 피해를 줄이고자 성공 여부 예측에 관한 다양한 연구가 진행되고 있다.

본 연구에서는 프로젝트 제목의 품사 출현 빈도 정보를 새로운 요인으로 설정하여 성공 여부 예측에 미치는 영향을 연구해보고자 한다. 이에 따라, 펀딩 목표 금액, 게시글의 이미지 개수 등과 같은 프로젝트 자체의 펀딩 정보만 고려한 모델과 해당 모델에 품사 정보를 적용한 모델을 비교하여 품사 정보가 프로젝트 성공 여부 예측에 미치는 영향에 대해 파악하고자 한다.

II. 데이터 수집 및 전처리

본 연구에서는 국내 최대 크라우드 펀딩 플랫폼인 와디즈(wadiz)에서 2019년 3월부터 2022년 8월까지 진행된 패션잡화 카테고리의 보상형 펀딩 프로젝트에 대해 크롤링을 진행하여 총 9,967개에 대한 프로젝트 정보를 수집하였다. 수집한 정보는 프로젝트에서 제공하는 정보인 펀딩 목표 금액, 게시글의 이미지 수, 제목 길이, 펀딩 기간, 펀딩 성공 여부이다. 또한, 수집한 데이터를 기반으로 메이커에 대한 신뢰성을 판단할 수 있는 지표를 추가 변수로 생성해주었다. 메이커가 진행한 프로젝트들의 경험과 성공 여부를 파악하기 위해 ‘성공’, ‘실패’, ‘신규’ 변수를 수집했다. ‘성공’, ‘실패’ 변수는 각각 이전 프로젝트의 성공, 실패 횟수를 나타낸다. ‘신규’ 변수의 경우, 이전 프로젝트 경험이 존재하면 0을, 존재하지 않은 경우는 1을 부여하였다. ‘pf’ 변수는 성공 여부를 나타내는 변수로 해당 프로젝트가 실패한 경우를 0을, 성공한 경우는 1을 부여하였다.

표 1. 입력 변수

Table 1. input variables

구분		변수명	설명
종속변수		pf	펀딩 성공 여부 (1 : 성공, 0 : 실패)
독립변수	펀딩 정보	sucess	펀딩 메이커의 이전 프로젝트까지 성공 횟수
		fail	펀딩 메이커의 이전 프로젝트까지 실패 횟수
		new	펀딩 메이커의 프로젝트 경험 여부 (1 : 신규메이커, 0 : 기존메이커)
		goal	펀딩 목표 금액
		img	펀딩 게시글의 이미지 수
		length	펀딩 게시글 제목 길이
		funding_term	펀딩 기간
	게시글 제목 품사 정보	Josa	조사
		Adjective	형용사
		Noun	명사
		Number	숫자
		Verb	동사

펀딩 게시글 제목의 품사별 빈도를 파악하기 위하여 대표적인 형태소 분석 및 품사 태깅 라이브러리 KoNLPy를 사용하였다. 본 연구에서는 KoNLPy

가 제공하는 Hannanum, KKma, Mecab, Komoran, Okt 형태소 분석기 중 Okt 형태소 분석기를 사용하여 품사 태깅을 진행하였다. 품사 태깅 결과로 얻은 ‘Noun’, ‘Josa’, ‘Verb’ 등 20개의 품사 정보 중 문장부호를 의미하는 ‘Punctuation’를 제외한 빈도 평균값 상위 5개 ‘Noun’, ‘Josa’, ‘Verb’, ‘Adjective’, ‘Numb’ 품사 정보만을 활용하여 분석하고자 한다. 해당 과정을 통해 선정한 변수들에 대한 자세한 설명은 [표1]과 같다. [3]

III. 데이터 분석 및 결과

본 연구의 목적인 품사의 영향을 파악하기 위하여 [표 1]의 펀딩 정보 변수만 훈련한 기본 모델과 기본 모델에 품사 정보 변수를 더한 모델을 구축하여, 두 모델의 성능을 비교해보고자 하였다.

훈련 데이터와 테스트 데이터를 7대 3의 비율로 나누어 LightGBM(이하 LGBM), GradientBoosting(이하 GBT), AdaBoost(이하 ADB)의 총 3가지 분류 기법을 활용하여 크라우드 펀딩 성공 여부 예측을 진행하였다. 모델의 성능 지표로는 혼동행렬의 accuracy, precision, recall, f1-score를 활용하였다. 각 모델의 성능은 [표2]와 같다.

표 2. 모델 성능 결과

Table 2. model performance result

모델	성능 지표	LGBM	GBT	ADB
펀딩 정보	accuracy	0.87	0.87	0.87
	precision	0.85	0.86	0.86
	recall	0.87	0.87	0.87
	f1-score	0.85	0.86	0.86
펀딩 정보 + 품사 정보	accuracy	0.87	0.88	0.88
	precision	0.86	0.87	0.87
	recall	0.87	0.88	0.88
	f1-score	0.86	0.86	0.87

LGBM 모델에서는 precision과 f1-score가, GBT 모델에서는 accuracy, precision과 recall이, ADB 모델에서는 모든 지표가 품사 정보를 더한 모델에서 0.01 높은 것으로 나타났다. 품사 정보를 더한 모델이 기본 모델보다 근소하지만, 더 나은 성능을 가지는 것을 확인하였다. 이에 따라 품사가 크라우드 펀딩 성공 여부 예측에 긍정적인 영향을 미치는 것을 알 수 있었다.

IV. 결론 및 시사점

본 논문에서는 품사 정보가 크라우드 펀딩 성공 예측에 미치는 영향에 대해 살펴보았다. 분석 결과, 품사 정보로 인하여 크라우드 펀딩 성공 예측 모델이 개선되는 것을 확인할 수 있었다. 본 연구는 크라우드 펀딩 성공 여부 예측에 대한 품사의 영향을 확인한 것에 의의가 있다. 그러나 이번 연구는 데이터 자체의 불균형으로 인하여 프로젝트 실패에 대한 예측이 어렵다는 한계점을 가지고 있다. 향후 진행되는 연구에서는 실패한 프로젝트에 대한 충분한 데이터를 확보하여 연구를 진행한다면 더 나은 결과를 도출할 수 있을 것이다.

참 고 문 헌

- [1] S. W. Choi, D. Y. Lee, et al., "An Empirical Study on Successful Crowdfunding", Asia-Pacific Journal of Business Venturing and Entrepreneurship Vol.12 No.2. pp.55-63.
- [2] D.J. Moon and S.H.Yoon, et al., "A Machine Learning Approach for the Success Prediction of Reward Crowdfunding Project", Korea Business Review, vol.24 No.3, pp.125-143, August 2020.
- [3] S. E. Park, J.Y. Kang, et al., "DS_077, The Complete Guide to Python Text Mining", wikibooks, pp. 36-42.