

멀티미디어 및 언어적 특성을 활용한 크라우드펀딩 캠페인의 성공 여부 예측

이강희[†], 이승훈^{**}, 김현철^{***}

Predicting Success of Crowdfunding Campaigns using Multimedia and Linguistic Features

Kang-hee Lee[†], Seung-hun Lee^{**}, Hyun-chul Kim^{***}

ABSTRACT

Crowdfunding has seen an enormous rise, becoming a new alternative funding source for emerging startup companies in recent years. Despite the huge success of crowdfunding, it has been reported that only around 40% of crowdfunding campaigns successfully raise the desired goal amount. The purpose of this study is to investigate key factors influencing successful fundraising on crowdfunding platforms. To this end, we mainly focus on contents of project campaigns, particularly their linguistic cues as well as multiple features extracted from project information and multimedia contents. We reveal which of these features are useful for predicting success of crowdfunding campaigns, and then build a predictive model based on those selected features. Our experimental results demonstrate that the built model predicts the success or failure of a crowdfunding campaign with 86.15% accuracy.

Key words: Crowdfunding, Kickstarter, Machine Learning, Deep Learning, Success Prediction

1. 서 론

크라우드 펀딩(Crowdfunding)은 프로젝트 아이디어를 광고함으로써 인터넷상의 사용자들이 온라인을 통해 후원하는 방법으로 지난 10년부터 자금 조달이 어려운 신생 기업이 투자를 받는 방법으로 떠오르고 있다. 크라우드 펀딩 산업의 규모는 2014년에 총 160억 달러가 투자되었고 2015년에는 344억 달러가 투자되면서 크라우드 펀딩 산업의 규모는 매년 두 배씩 급증하고 있다[1].

크라우드 펀딩의 투자 모델에는 4가지 유형이 있다[2]. 첫 번째 모델은 기부 모델(Donation-based)이

다. 기부 모델은 투자자가 어떠한 보상을 전제로 하지 않는 기부형식의 투자이다. 두 번째는 대출 모델(Lending-based)로서 투자자의 자금이 대출의 형태로 제공되며 이자를 통해 일부 수익률을 기대한다. 세 번째는 지분 투자 모델(Equity-based)로서 주식 또는 수익 증권을 가지는 방식이다. 마지막으로, 보상 모델(Reward-based)은 투자자가 투자에 대한 보상을 받는 모델이다. 본 연구에서는 4가지 모델 중 가장 널리 퍼진 보상 모델을 연구한다. 2016년 1월을 기준으로 가장 큰 규모의 보상 모델 크라우드 펀딩 사이트는 킥스타터(kickstarter.com)로 266만 건의 프로젝트에 20억 달러 이상을 모금했다[3].

* Corresponding Author: Hyun-chul Kim, Address: (31066) 31, Sangmyeongdae-gil, Dongnam-gu, Cheonan-si, Chungcheongnam-do, Korea, TEL: +82-41-550-5403, FAX: +82-41-550-5369, E-mail: hkim@smu.ac.kr

Receipt date: Jan. 12, 2018, Approval date: Feb. 7, 2018

[†] Department of Software., Sangmyung university
(E-mail: lkh7054@gmail.com)

^{**} Department of Software., Sangmyung university
(E-mail: mr.leesh90@gmail.com)

^{***} Department of Software., Sangmyung university

* This research was supported by a 2016 Research Grant from Sangmyung University.

키스타터에서 펀딩 기간과 목표 금액은 프로젝트 제작자에 의해 정해지며, 펀딩에 성공하는 조건은 정해진 기간이 종료됐을 때를 기준으로 후원된 금액이 목표 금액을 넘는 것이다. 펀딩이 성공하면 프로젝트에 후원된 금액이 프로젝트 제작자에게 전달되고 후원 금액별 보상이 후원자들에게 전달되는 것을 약속한다. 만약 펀딩이 실패한다면 후원된 금액은 다시 후원자들에게 반환된다.

키스타터에서 펀딩에 성공하는 프로젝트는 전체 프로젝트의 44%에 불과하다[4]. 펀딩에 실패하면 프로젝트 제작자는 투자를 받을 수 없기 때문에 프로젝트 제작자는 프로젝트의 성공 가능성을 사전에 파악하여 그에 따라 대응할 수 있어야 한다. 프로젝트의 성공 예측 결과가 실패로 나온 제작자는 프로젝트의 캠페인을 수정하거나 SNS를 통해 프로젝트를 광고하여 펀딩에 성공하도록 만들 수 있고, 성공 예측 결과가 성공으로 나온 프로젝트는 자신의 프로젝트를 더욱 홍보하거나 확장하는 방법을 검토할 수 있다. 키스타터의 성공 예측은 후원자에게도 유용하다. 펀딩에 실패한 프로젝트일 경우 후원자는 후원한 금액을 돌려받지만 프로젝트를 알아보고 투자하는데 사용된 시간이 낭비되며 관심 있는 프로젝트일 경우 후원자가 적극적인 홍보하는 것에 동기부여를 할 수 있다.

본 연구에서는 키스타터에 등록된 프로젝트 중 가장 활발하게 펀딩이 진행되고 실용적인 요구사항을 충족시키는 대표적인 카테고리인 Technology 카테고리 and 예술적인 아름다움을 강조하는 Art 카테고리들의 프로젝트 정보를 수집하고, 인공지능과 멀티미디어 특성 및 언어적인 특성 정보를 이용하여 각 프로젝트들의 성공 여부를 예측한다.

2. 관련 연구

Mollick[5]은 키스타터에서 성공과 실패에 대한 연구 결과를 제공한다. 그는 성공을 위한 요인에 대해서 프로젝트의 품질과 크라우드 펀딩의 지리적 영향 등 다양한 통계를 제공한다. 소셜 네트워크(Facebook)와 프로젝트 품질(프로젝트 설명, 맞춤법 오류, 비디오)은 프로젝트 성공의 긍정적 요인이라는 것을 발견했다. Greenberg[6]는 실패한 프로젝트 중 재도전하는 프로젝트는 2% 미만이며 재도전한 프로젝트의 43%만이 프로젝트를 성공시키는 것을 발견

했다. 실패한 프로젝트가 재도전할 때 실패한 경험 중 피드백을 받는 과정이나 새로운 소셜 네트워크에 대한 내용을 강조할 것을 추천했다.

Greenberg의 연구에서는[7] 프로젝트를 시작할 때 사용하는 정보만으로 키스타터의 프로젝트의 성공을 Random Forest 알고리즘을 사용해 68%의 정확도로 예측했다. Etter는[4] Greenberg의 연구[7]를 바탕으로 Sidekick[8]과 같은 프로젝트의 성공 확률을 예측하는 도구를 만들었다. Chen[9]은 프로젝트 정보와 프로젝트와 연관된 외부 미디어의 정보를 사용해 SVM(Support Vector Machine) 알고리즘을 이용하여 67%의 정확도로 예측했다. 그리고 외부 미디어인 YouTube와 Twitter가 프로젝트의 성공에 영향을 미치지 못한다는 것을 발견했다. Desai의 연구에서는[10] 모든 카테고리에서 성공한 프로젝트 7,862개와 실패한 프로젝트 18,681개의 정보를 수집하고 LIWC[11, 12]와 Stanford CoreNLP[13]를 이용하여 키스타터의 언어적 특징을 분석하고 프로젝트의 정보와 언어적 특징의 정보를 이용하여 Naive Bayes, Decision Tree, Logistic Regression, SVM 알고리즘을 이용해서 프로젝트의 성공 예측을 하였다.

위의 연구에서는 키스타터 프로젝트의 성공 예측에 프로젝트의 정보와 멀티미디어 정보 또는 프로젝트의 정보와 언어적 특징과 같이 일부분만 조합하여 사용했지만 본 연구에서는 프로젝트의 정보, 언어적 특징 정보, 제작자 또는 후원자의 정보와 멀티미디어 정보를 모두 조합하여 사용했고 지금까지 사용하지 않았던 심층 학습 알고리즘인 Multi-layer Perceptron을 사용해 프로젝트의 성공을 예측했다. 본 연구에서는 86.15%의 정확도로 위의 연구들의 보다 더 정확하게 프로젝트의 성공을 예측할 수 있었다.

3. 방법론

3.1 데이터 세트

Fig. 1은 키스타터 프로젝트의 캠페인의 예이다. 키스타터에서 프로젝트의 구성은 1) 캠페인 섹션, 2) 업데이트 섹션, 3) 코멘트 섹션, 4) 커뮤니티 섹션, 5) FAQ 섹션이 있다. 캠페인 섹션은 프로젝트 제작자가 이미지 또는 비디오 등을 이용하여 프로젝트 아이디어를 소개하고 설명한다. 업데이트 섹션은 프

로젝트 제작자가 프로젝트 진행 상황을 게시하며, 코멘트 섹션은 후원자와 작성자 모두 자유롭게 의견과 피드백을 남길 수 있다. 커뮤니티 섹션은 후원자 목록이 포함되어 있으며, FAQ 섹션은 캠페인에 대한 후원자의 자주 묻는 질문 목록과 제작자의 답변이 있다.

킵스타터에는 Art, Comic, Food, Design, Technology 등 15가지의 카테고리가 있다. Dey의 연구에서는[14] 실용성과 예술성에 대해 세 카테고리(Technology, Design, Fashion)가 다른 강조점을 가진다고 설명한다. 본 연구에서는 여러 카테고리 중 실용적인 요구사항을 충족시키는 Technology와 예술적인 아름다움을 강조하는 Art를 선정하고, 2개의 카테고리에서 성공한 프로젝트와 실패한 프로젝트를 1,000개씩 수집하여 총 4,000개의 프로젝트 데이터를 수집했다.

3.2 Features

데이터 세트는 i) 프로젝트의 기본 정보, ii) 멀티미디어 정보, iii) 프로젝트 제작자의 정보, iv) 후원자의 정보, v) 언어적 특징으로 구성된다.

i) 프로젝트 기본 정보

킵스타터 홈페이지에서 제공하는 프로젝트의 기본 정보이다. 1) 펀딩에 성공하기 위한 목표 금액, 2)

프로젝트를 진행하면서 제작자가 올리는 업데이트의 수, 3) 제작자 또는 후원자가 작성하는 코멘트의 수, 4) 전체 후원자의 수, 5) 보상을 원하는 후원자의 수, 6) 보상의 종류가 해당된다.

ii) 멀티미디어 정보

프로젝트 제작자가 킵스타터에서 프로젝트를 소개하기 위해 사용했던 멀티미디어의 정보이다. 1) 메인 비디오의 유무, 2) 캠페인, 업데이트 섹션에 삽입된 비디오의 수, 3) 캠페인, 업데이트 섹션에 사용된 이미지의 수를 멀티미디어 정보로 사용한다.

iii) 프로젝트 제작자의 정보

프로젝트 제작자가 공개한 자신의 정보와 킵스타터에서 했던 활동과 관련된 정보이다. 1) LinkedIn이나 Twitter와 같은 외부 링크 또는 웹 사이트의 수, 2) Facebook ID의 존재 여부, 3), 제작자가 후원한 프로젝트의 수, 4) 지금까지 진행한 프로젝트의 수, 5) 현재 프로젝트를 시작하기 전에 다른 프로젝트에 남긴 코멘트의 수, 6) 현재 프로젝트를 시작한 후에 다른 프로젝트에 남긴 코멘트의 수, 7) 제작자의 계정이 생성되고 프로젝트를 시작하기까지의 시간을 제작자의 정보로 사용한다.

iv) 후원자의 정보

킵스타터에서 후원자의 활동과 관련된 정보이다.

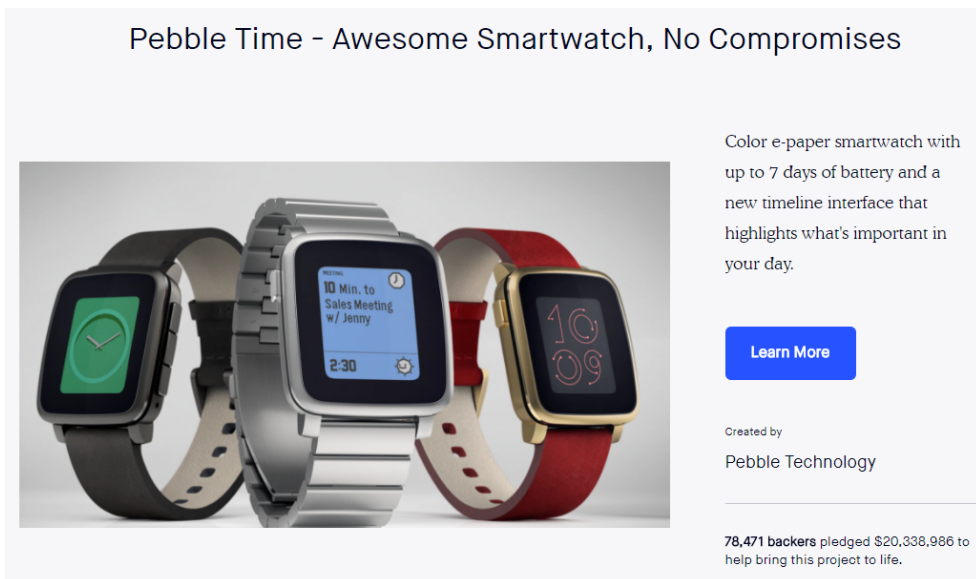


Fig. 1. A Kickstarter Project.

1) 처음으로 투자하는 후원자의 수, 2) 투자한 경험이 있는 후원자의 수가 해당된다.

v) 언어적 특징

언어는 현실에 대한 자신의 생각, 감정 또는 직관을 식별할 때 중요한 매개 변수이다. 본 연구에서는 프로젝트의 캠페인, 업데이트, 코멘트 섹션에 있는 텍스트 정보를 분석하여 언어적 특징을 추출했다. 언어적 특징을 추출하는 방법으로 NLTK[15], LIWC [11, 12], Stanford CoreNLP[13]를 사용했다.

- 양 : 단어, 문자, 동사, 명사, 형용사, 부사, 문장, 구, 명사 구, 절, 구두점, 쉼표의 수
- 복잡성 : 절의 평균 길이, 문장의 평균 길이, 단어의 평균 길이
- 불확실성 : 조동사, 2인칭 대명사, 3인칭 대명사의 수

- 다중성 : 기능적인 단어의 수(예 : 전치사, 접속사, 관사)
- 비직접성 : 1인칭 단수 대명사, 1인칭 복수 대명사 수
- 표현력 : (형용사 수 + 부사 수) / (명사 수 + 동사 수)
- 다양성 : 어휘의 다양성 (다른 단어의 수 / 단어의 수), 중복된 어휘의 비율 (기능 단어의 수 / 문장의 수)
- 비공식성 : 철자가 틀린 단어의 수, 문장에서 철자가 틀린 단어의 비율
- 사회성 정보 : E-mail, URL, 전화번호를 언급한 횟수
- 상대성 : 위치, 조직, 사람, 시간, 과거 시제, 현재 시제, 미래 시제를 나타내는 단어의 수
- 인지 과정 : 긍정/부정적인 감정 단어, 인과 관계, 확신, 생각, 부정, 행동, 추측, 차별, 배타적인 단어

Table 1. CFS results and the descriptive statistics of CFS results

Feature	Mean(SD)		p-value
	success	fail	
Number of projects(backed by the creator)	9.37(33.48)	1.76(6.98)	***
Number of updates	12.31(12.25)	1.21(2.96)	***
Number of backer comments	124.99(696.4)	1.19(7.11)	***
Number of backers	424.26(2681.5)	15.36(45.05)	***
Number of reward backers	402.68(2421.66)	13.32(43.19)	***
Number of update images / Number of updates	0.82(1.31)	0.18(0.88)	***
(Campaign)Number of function words / Number of sentences	5.46(2.42)	6.15(4.23)	***
(Campaign)Number of cognitive process words	8.41(2.62)	9.14(3.4)	***
(Update)Number of first plural person words / Number of words	0.017(0.015)	0.006(0.014)	***
(Update)Number of third person words / Number of words	0.012(0.008)	0.004(0.009)	***
(Update)Number of location words / Number of updates	0.64(1.16)	0.13(0.78)	***
(Update)Number of location words	7.43(15.86)	0.45(2.53)	***
(Update)Number of person words	1.02(1.44)	0.28(1.15)	***
(Update)Number of time words	33.13(55.48)	1.67(6.35)	***
(Update)Number of insight words	1.27(0.82)	0.56(1.25)	***
(Update)Number of tentative words	1.59(1.15)	0.47(1.25)	***
(Comment)Number of email words / Number of comments	0.01(0.07)	0.0004(0.007)	***

Significance: *** p < 0.001, ** p<0.01, * p<0.05, If p-value is less than 0.05 then it becomes more significant for the model, SD: standard deviation

의 수

• 가독성 : 음절, 단어 및 문장 수를 계산하여 텍스트의 가독성을 평가(ARI[16], CL[17], GF[18], FKGL[19], FRES[19])

4. 결 과

4.1 Feature Selection

Feature Selection은 기계 학습의 전처리 단계로서 더 높은 학습 정확도를 위해 Feature의 하위 집합을 선택하는 과정이다. 본 연구에서는 실용적이고 분류 정확도와 효율성 측면에서 다른 필터보다 우수한 효과를 가지고 있는 CFS 필터를 사용하며 Greedy

검색 방법보다 높은 분류 정확도를 제공하는 Best First 검색 방법을 조합하여 Feature의 하위 집합을 선별한다[20, 21].

BestFirst 검색 방법을 사용한 CFS 알고리즘을 통해서 259개의 Feature 중 17개의 Feature를 선별했다. Table 1은 CFS 알고리즘을 통해서 선별된 17개의 Feature이다.

프로젝트에 성공하는 제작자의 경우 실패하는 제작자보다 평균적으로 5.32배 더 많은 다른 프로젝트에 투자하며 업데이트 또한 10.17배 정도 많이 한다. 성공한 프로젝트의 경우 업데이트 1회당 평균적으로 사용되는 사진의 수가 4.61배 차이 났으며 위치, 사람, 시간, 생각, 추측을 뜻하는 단어가 실패한 프로젝트

Table 2. Selected feature set

Group	Section	Feature
Project Information		Number of updates
		Number of backer comments
		Number of backers
		Number of reward backers
Creator Information		Number of projects(backed by the creator)
Multimedia Information	Campaign	Presence of main video
		Number of campaign videos
		Number of campaign images
		Number of campaign audios
	Update	Number of update videos
		Number of update videos / Number of updates
		Number of update images
		Number of update images / Number of updates
		Number of update audios
		Number of update audios / Number of updates
Linguistics cues	Campaign	Number of function words / Number of sentences
		Number of cognitive process words
	Update	Number of first plural person words / Number of words
		Number of third person words / Number of words
		Number of location words / Number of updates
		Number of location words
		Number of person words
		Number of time words
		Number of insight words
	Number of tentative words	
Comment	Number of email words / Number of comments	

트보다 많이 사용되는 것을 알 수 있었다. 반대로 실패한 프로젝트는 성공한 프로젝트보다 캠페인 섹션에서 중복되는 단어나 “cause”, “know”, “ought”와 같은 인지 과정과 연관 있는 단어가 더 많이 사용되는 것을 알 수 있었다.

CFS 알고리즘의 결과로 추출된 Feature들을 T-test를 이용하여 p-value를 확인했다.

본 연구에서는 CFS 알고리즘을 통해서 선별된 17개의 Feature와 멀티미디어 Feature 9개를 추가하여 Table 2와 같이 학습 데이터 세트를 만들었다.

4.2 성공 예측

본 연구에서는 WEKA[22]를 사용하여 SVM(Support Vector Machine), LR(Logistic Regression)과 같은 2가지 기계 학습 알고리즘과 심층 학습에서 사용되는 MLP(Multi-layer Perceptron)를 이용하여 성공 여부 분류 모델을 만들고 실험하였다.

각 모델의 파라미터 설정에서 Batch Size는 WEKA의 기본값인 100으로 설정하고 SVM은 커널 함수를 이용하여 선형 공간을 비선형 공간으로 변화시켜서 선형 분류가 불가능한 데이터를 선형 분류가 가능하게 한다. 커널 함수로는 선형 커널, 가우시안 커널, 다항식 커널 등이 있다[23]. 본 연구에서는 SVM의 커널 함수로 다항식 커널(PolyKernel)을 사용했다. MLP는 히든 레이어의 구성을 다수의 레이어와 노드로 구성할 수 있다. 본 연구에서는 히든 레이어의 개수를 2개로 설정하며 첫 번째 히든 레이어의 노드를

100개, 두 번째 히든 레이어의 노드를 200개로 설정했다.

Table 3에 나타난 결과를 보면 LR은 86.15%의 확률로 프로젝트의 성공 여부를 예측하는 3가지 알고리즘 중에서 가장 높은 결과를 보였으며 성공 프로젝트 2,000개 중 1,649개와 실패 프로젝트 2,000개 중 1,797개를 분류하였다. MLP은 두 번째로 높은 정확도인 83.87%로 성공 여부를 예측한다. 실험을 한 모든 알고리즘 중 SVM은 가장 낮은 정확도인 81.95%로 성공 여부를 예측했다.

4.3 Feature 그룹별 성공 예측

Feature Selection에서 선별한 26개의 Feature를 그룹별로 분류하여 성공 예측 실험을 해본 결과 Table 4과 같은 결과를 얻을 수 있다.

프로젝트 정보에 해당하는 Feature들만 사용했을 때 LR 알고리즘에서 Table 3의 결과와 1.4%의 차이로 비슷한 성능을 보였고 MLP 알고리즘에서 3.2%의 차이를 보였다. 언어적 특징의 Feature들은 두 번째로 뛰어난 성능을 보였으며 프로젝트 제작자 정보 Feature들은 65.37%로 가장 낮은 성능을 보여주는 그룹이다.

5. 결론

클라우드 펀딩의 규모가 점점 성장하면서 벤처 기업의 투자 대안으로 더욱 강조되고 있다. 본 연구에서는 보상형 클라우드 펀딩의 기업 중 최대 규모의

Table 3. Performance comparison of different classification algorithms

Algorithm	Precision	Recall	Accuracy(%)	AUC
LR	0.824	0.89	86.15	0.965
SVM	0.839	0.807	81.95	0.815
MLP	0.805	0.862	83.87	0.679

Table 4. Prediction accuracy by feature group

Group	Accuracy(%)		
	LR	SMO	MLP
Project Information	84.75	78.65	82.95
Multimedia Information	73.35	68.35	71.53
Creator Information	65.37	50.97	62.55
Linguistics cues	82.77	80.67	82.08

기업인 킥스타터에서 Technology와 Art 카테고리를 대상으로 성공했거나 실패한 프로젝트의 데이터를 수집하였고 프로젝트의 정보, 멀티미디어 정보, 프로젝트 제작자의 정보, 후원자의 정보, 언어적 특징을 분석하여 킥스타터의 성공 예측에 필요한 Feature들을 발견했으며 프로젝트 정보에 해당하는 Feature가 성공을 예측하는데 효과적이라는 것을 발견했다. 본 연구에서는 3가지의 기계 학습 알고리즘을 이용해서 킥스타터의 성공을 예측하는 다른 연구보다 더 높은 정확도인 86.15%의 확률로 프로젝트의 성공을 예측하였다.

REFERENCE

- [1] Massolution, 2015CF Crowdfunding Industry Report, <http://reports.crowdsourcing.org>, (accessed Jun., 1, 2018).
- [2] D.J. Cumming, G. Leboeuf, and A. Schwienbacher, *Crowdfunding Models: Keep-it-all vs. All-or-nothing*, Call for Papers : Paris December 2014 Finance Meeting Eurofidai-affi Paper, 2014.
- [3] The Funding Pledged to Kickstarter Projects, <https://www.statista.com/statistics/310218/total-kickstarter-funding/>, (accessed Jun., 1, 2018).
- [4] V. Etter, M. Grossglauser, and P. Thiran, "Launch Hard or Go Home!: Predicting the Success of Kickstarter Campaigns," *Proceeding of the Association for Computing Machinery Conference on Online Social Networks*, pp. 177-182, 2013.
- [5] E. Mollick, "The Dynamics of Crowdfunding: An Exploratory Study," *Journal of Business Venturing*, Vol. 29, No. 1, pp. 1-16, 2014.
- [6] M.D. Greenberg and M.G. Elizabeth, "Learning to Fail: Experiencing Public Failure Online through Crowdfunding," *Proceeding of the Association for Computing Machinery Special Interest Group on Computer-human Interaction Conference on Human Factors in Computing Systems*, pp. 581-590, 2014.
- [7] M.D. Greenberg, B. Pardo, K. Hariharan, and E. Gerber, "Crowdfunding Support Tools: Predicting Success and Failure," *Proceeding of the Association for Computing Machinery Special Interest Group on Computer-human Interaction Conference on Extended Abstracts on Human Factors in Computing Systems*, pp. 1815-1820, 2013.
- [8] Kickstarter Success Prediction Service, <http://sidekick.epfl.ch/>, (accessed Jun., 1, 2018).
- [9] K. Chen, B. Jones, I. Kim, and B. Schlamp, *KickPredict: Predicting Kickstarter Success*, Technical Report, California Institute of Technology, 2013.
- [10] N. Desai, R. Gupta, and K. Truong, *Plead or Pitch? The Role of Language in Kickstarter Project Success*, Technical Report, Stanford University, 2015.
- [11] J.W. Pennebaker, R.L. Boyd, K. Jordan, and K. Blackburn, The Development and Psychometric Properties of LIWC, Technical Report <https://repositories.lib.utexas.edu/handle/2152/31333>, (accessed Jun, 1, 2018).
- [12] J.W. Pennebaker, M.E. Francis, and R.J. Booth, "Linguistic Inquiry and Word Count: LIWC 2001," *Journal of Mahway: Lawrence Erlbaum Associates*, Vol. 71, No. 2001, pp. 2001, 2001.
- [13] C.D. Manning, M. Surdeanu, J. Bauer, J.R. Finkel, S. Bethard, and D. McClosky, "The Stanford CoreNlp Natural Language Processing Toolkit," *Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pp. 55-60, 2014.
- [14] S. Dey, B. Duff, K. Karahalios, and W.T. Fu, "The Art and Science of Persuasion: Not All Crowdfunding Campaign Videos Are The Same," *Proceeding of the Association for Computing Machinery Conference on Computer Supported Cooperative Work and Social Computing*, pp. 755-769, 2017.
- [15] E. Loper and S. Bird, "NLTK: The Natural Language Toolkit," *Proceeding of the Association for Computational Linguistics-02*

Workshop on Effective Tools and Methodologies for Teaching Natural Language Processing and Computational Linguistics-volume 1, pp. 63-70, 2002.

- [16] R.J. Senter and E.A. Smith, *Automated Readability Index*, Technical Report, University of Cincinnati, 1967.
- [17] M. Coleman and T.L. Liao, "A Computer Readability Formula Designed for Machine Scoring," *Journal of Applied Psychology*, Vol. 60, No. 2, pp. 283, 1975.
- [18] R. Gunning, *The Technique of Clear Writing*, McGraw-Hill Publisher, New York, 1952.
- [19] R. Flesch, "A New Readability Yardstick," *Journal of Applied Psychology*, Vol. 32, No. 3, pp. 221-233, 1948.
- [20] N. Williams and S. Zander, *Evaluating Machine Learning Algorithms for Automated Network Application Identification*, Technical Report, Swinburne University of Technology, Centre for Advanced Internet Architectures, 2006.
- [21] H.C. Kim, K.C. Claffy, M. Fomenkov, D. Barman, M. Faloutsos, and K.Y. Lee, "Internet Traffic Classification Demystified: Myths, Caveats, and the Best Practices," *Proceeding of the Association for Computing Machinery International Conference on Emerging Networking Experiments and Technologies*, pp. 11, 2008.
- [22] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I.H. Witten, "The WEKA Data Mining Software: an Update," *Journal of The Association for Computing Machinery Special Interest Group on Knowledge Discovery and Data Mining Explorations Newsletter*, Vol. 11, No. 1, pp. 10-18, 2009.

- [23] H. Lee, "Use of the Moving Average of the Current Weather Data for the Solar Power Generation Amount Prediction," *Journal of Korea Multimedia Society*, Vol. 19, No. 8, pp. 1530-1537, 2016.



이 강 희

2012년 3월~현재 상명대학교 컴퓨터소프트웨어공학과 학사과정
관심분야: 데이터 과학



이 승 훈

2015년 2월 상명대학교 컴퓨터소프트웨어공학과 졸업
2017년 2월 상명대학교 컴퓨터공학과 공학석사
관심분야: 데이터 마이닝



김 현 철

1995년 2월 KAIST 전산학과 졸업
1997년 2월 KAIST 전산학과 공학 석사
2005년 2월 KAIST 전자전산학과 공학박사
2006년 1월~2007년 12월 CAIDA (Center for Applied Internet Data Analysis), San Diego Supercomputer Center 박사 후 방문 연구원
2008년 3월~2012년 2월 서울대학교 컴퓨터공학부 BK 조교수
2012년 3월~현재 상명대학교 소프트웨어학과 조교수
관심분야: 데이터 과학, 네트워크/인터넷 과학