Proceedings of the 3rd International Conference on Signal Processing and Machine Learning
DOI: 10.54254/2755-2721/5/20230670

Creating a computer mediated equity crowdfunding mechanism based on regression and sentiment analysis

Weiwei Wang

Trinity College of Arts and Sciences, Duke University, Durham, NC, 27708, U.S.A.

631301120207@mails.cqjtu.edu.cn

Abstract. The rise of crowdfunding provides a viable alternative to the traditionally opaque capital market and significantly reduces the cost due to information asymmetry among founders, investors, and intermediaries. Numerous research studies have been conducted to predict the success of a crowdfunding campaign, including subjectivity vs objectivity, overstatement, keywords, frequency, length of interactions between founders and investors, etc. However, very few of them, if any, shed lights on actionable implementations to offer suggestions and aid entrepreneurs, especially those from disadvantaged backgrounds and don't have access to financial advisors, navigating through the fundraising process. Through analysing two studies examining founder and investor communication, one focusing on unilateral communication and message conveyed by founders to potential investors, the other focusing on bilateral interaction between the two parties, using sentiment analysis, this paper aims to provide a reference for creating a computer-mediated mechanism, offering entrepreneurs suggestions for future campaigns, and bringing equity to the capital market across different geographies.

Keywords: equity crowdfunding, speech recognition, sentiment analysis, human-computer interaction, machine learning.

1. Introduction

With the rise of Sharing Economy and the Internet, crowdfunding platforms (i.e., Kickstarter, Indiegogo) have gained growing traction among entrepreneurs and investors [1]. By Sep 2022, Kickstarter alone has facilitated US$ 6.8bn funding pledged to its projects with a success rate of 40% [2,3]. There are various types of crowdfunding models based on the nature of financing, namely debt financing, equity financing and charity financing [4]. Considering the investors for debt and donation crowdfunding rounds are not necessarily looking for pure investment return, the focus of this paper is equity crowdfunding.

Crowdfunding platform provides a viable alternative to the traditional fundraising model, in which investment banks and financial advisors tend to charge a hefty fee in turn for introductions to renowned institutions and preparation for pitching materials. Hence this study aims to democratize access to capital for marginalized entrepreneurs who otherwise do not have access to a network or financial advisor in a traditional fundraising process. Despite all benefits, how to effectively present the project and appeal to potential investors within the framework on different platforms could be challenging for those entrepreneurs who may not be experienced with the pitching process. Platforms such as Kickstarter allow entrepreneurs to include semantic information such as project overview and FAQ in the campaign,

DOI: 10.54254/2755-2721/5/20230670

along with a supplementary video. With the rise in volume and popularity, numerous studies have examined properties that constitute successful projects and attempted to predict funding outcomes based on various components of a campaign. For instance, project duration and interaction between entrepreneurs and potential investors (i.e., quantity, sentiment, length) [5].

In addition, it is also worth noting the two different types of information flow during a crowdfunding campaign: unilateral communication (i.e., project description, team member, video pitch, etc.), which offers investors information about the project. And bilateral communication (i.e., comments and updates), which refers to an interaction between founders and investors. Despite a school of research and literature around the former, components of the latter, including the frequency, length, sentiment of updates, and conversation, have not been well examined [6-8].

The object of the study is to evaluate various dimensions of a campaign to predict the outcome of an equity crowdfunding campaign and to provide a reference for the design of a computer-mediated mechanism to facilitate the equity crowdfunding process.

2. Method and results

2.1. Unilateral communication and source data overview

Unilateral communication refers to the one-way campaign materials prepared by the founders. In this study, the effect of the video component was explored using a random sample from Kickstarter. The dataset contains 148,398 campaigns that took place between 2014/3/1 and 2016/2/29. Excluding campaigns whose dataset is not comparable, terminated, or involves intellectual property breaches, the final sample contains 500 campaigns with an average pledged amount of US$11,522 from 169 investors per campaign [6].

2.1.1. Speech recognition process. Key success characteristics and variables, including sentiment-related metrics from the project description and supplementary video, are listed in Table 1 below. Watson Tone Analyzer by IBM (sample UI shown in Figure 1) was used to categorize and assess sentiment, linguistic styles, and social characteristics [6]:

1.Sentiment: Joy, Anger, Sadness, Disgust, Fear.

2.Linguistic Style: Analytical, Confident, Tentative.

3.Social Characteristics: Agreeableness, Extraversion, Openness, Emotional Range, Conscientiousness.

According to Kim et al., the speech was defined as a continuous auditory information stream with some degree of probability for different interpretations. On a high level, to render the stream of audio input into an analyzable format, the information flow is first split into a smaller unit of individual utterances, such as phonemes. The processed dataset then goes into the Long Short-term Memory Recurrent Neural Networks (LSTM RNNs) model, which will best fit the raw data with the sample sentiments list.

DOI: 10.54254/2755-2721/5/20230670

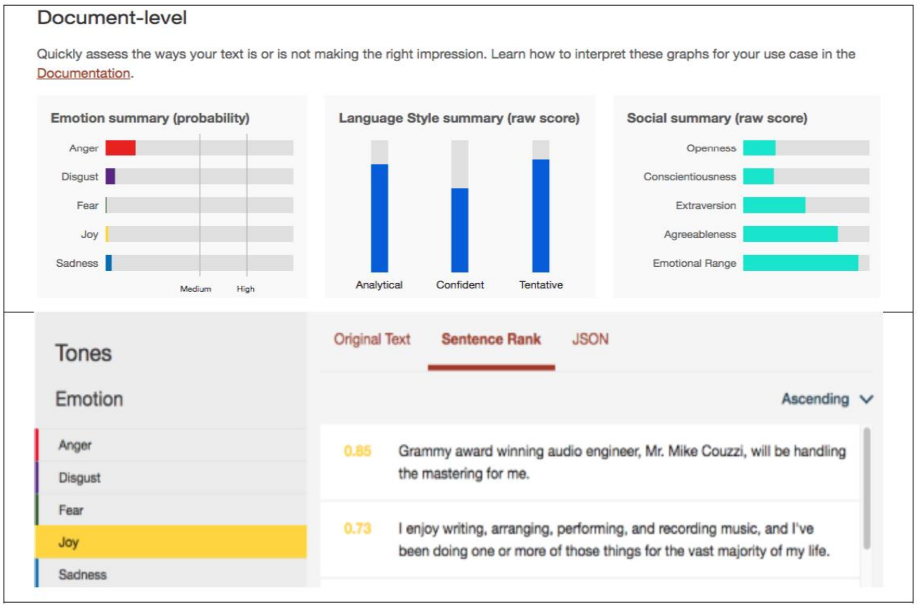

Figure 1. Sample UI of IBM Watson Tone Analyzer [6].

Table 1. Variable description of Project and Founding Team [6].

Variable Type	Control Variables	Descriptions
Project Features	fundraising_goal	The minimum amount of funding the project founders aim to raise
	description_length	Project description word count
	risk_description_length	Project opportunity and risk section word count
	fundraising_duration	Fundraising campaign duration (days)
	number_photo	Number of photos posted in the campaign
	number_video	Number of videos posted in the campaign
	length_video	Length of video in seconds
	staff_pick	If the platform has featured the campaign in the recommendation sector
	project_social_exposure	The number of external links such as personal website, Facebook, Twitter, etc.
	num_rewards	Number of rewards in the project
Founder Features	investment_history	Number of projects that the founder has invested in
	startup_history	Number of projects that the founder-led before
	facebook_connected	If the founder has shared on their own social media (i.e., Facebook)
	gender	Founder's predicted gender based on a given name

2.1.2. Preliminary results and discussion. Since the pledge amounts across the 500 projects were highly skewed, the Poisson model was used as the base of analysis, and variables were added to estimate parameters progressively. As shown in Table 2, only project variables were included for model 1. In model 2, text narratives were added to the model and proved to be a vital indicator of the total pledged amounts. For instance, fearful descriptions negatively impacted the outcome. In model 3, speech styles were included in the model along with control variables, showing that analytical speaking exhibits a strong positive impact. To sum up, that speech and text style can serve as references, and predictor for crowdfunding outcome is supported. Please note that t statistics are shown underneath the p-value.

Table 2. Correlation between fundraising outcome and speech/text styles [6].

	Model	**#1**	**#2**	**#3**	**#4**
Type	**Variables**	**Applied**	**Applied**	**Applied**	**Applied**
Speech	Analytic			**0.202****	**0.213*****
				(2.94)	**(3.56)**
	Extrovert			0.091	0.173
				(1.89)	(1.92)
	Confident			**0.143****	0.099
				(2.80)	(1.51)
	Empathic			-0.122	-0.197
				(1.02)	(1.86)
Text Narratives	Analytic		-0.015		-0.116
			(-0.21)		(-0.24)
	Extrovert		-0.163		-0.165
			(-1.60)		(-1.72)
	Mournful		-0.022		0.012
			(-1.10)		(0.15)
	Funny		0.148		0.131
			(1.02)		(1.16)
	Fearful		**-0.390****		**-0.374****
			(-2.60)		**(-2.89)**
	Project Characteristics	Yes	Yes	Yes	Yes
	Founder Characteristics	Yes	Yes	Yes	Yes
	Category Control Variables	Yes	Yes	Yes	Yes
	Observations	500	500	500	500
	Pseudo R2	0.401	0.433	0.473	0.520

Significance Level:
*p<0.05, **p<0.01, ***p<0.001

2.2. Bilateral communication and source data overview

Despite having a similar profile, geography, category, and team background, two crowdfunding projects may have drastically different outcomes – interaction between founders and investors is the missing variable in the equation. Communications facilitated by the platform and computer offer a desirable way for founders to help investors better understand the projects and for investors to build conviction in investing eventually [9-12]. Founders can either post project updates regarding campaign status in the "updates" section or reply to questions and complaints from investors in the Q&A section.

In this study, the effects of comment quality, sentiment, and length, as well as the ratio, length, and speed of reply on fundraising outcome, were examined. The 959 Chinese crowdfunding projects are from Dreamore, and sentiment analysis was conducted using the BosonNLP algorithm. Launched in Sep 2011, Dreamore, a Chinese crowdfunding platform, operates under a similar model as Kickstarters and Indiegogo. 393 projects were successful, and the total pledges of all 959 projects were CNY 20,041, 884 (equivalent to US$ 2,811,442) from 54,262 investors. Full dataset overview is shown in Table 3.

DOI: 10.54254/2755-2721/5/20230670

Table 3. Key Project Variable Overview [5].

Variable	Minimum	Maximum	Average	Std. Dev.
Success	0	1	0.41	0.49
Fundraising Amount (CNY)	100	800,000	35,406.63	87,578.65
Duration (Day)	1	112	35.19	14.57
Video	N/A	5	0.38	0.52
Picture	N/A	69	12.05	8.80
Update	1	16	1.05	0.57
Experience as Founder	1	11	1.78	2.32
Experience as Investor	N/A	20	0.49	1.62
Comment Quantity	N/A	581	5.98	21.24
Comment Sentiment	0.02	1	0.79	0.19
Comment Length (Byte)	N/A	481	22.96	27.39
Reply Ratio	N/A	1.5	0.07	0.21
Reply Length (Byte)	N/A	102	5.76	14.49
Reply Speed (Hour)	N/A	195	1.07	7.98

2.2.1. Research framework and methodology. The research framework is shown below in Figure 2, and an overview of the variables is in Table 5. Classic variables – including project characteristics, past experience of the founders, and investors that have been verified in the past literature as the determinants of project success – were deemed as control variables in the study.

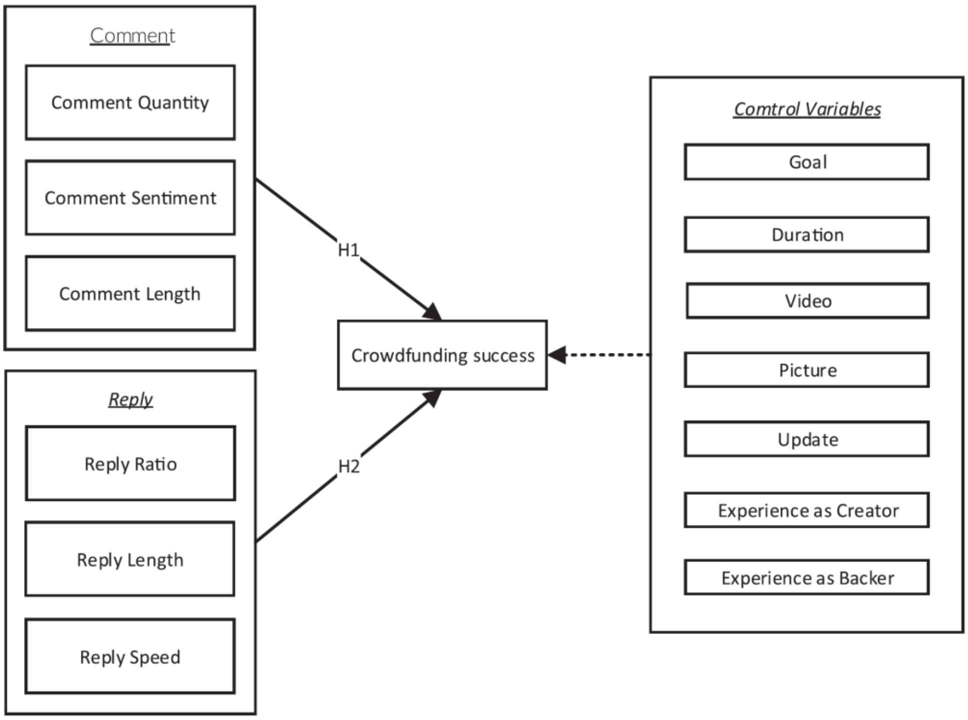

Figure 2. Research framework diagram [5].

Table 4. Comment sentiment analysis example [5].

	Positive comment	Negative comment
Input Comment	I really enjoyed my experience and can tell the seller put a lot of thought into the process. The video is very informative as well.	There is a very strong unpleasant feeling.
Comment Sentiment	[0.985, 0.015]	[0.004, 0.996]

Table 5. Description of Project Variables [5].

Variable Type		**Description**
Depend. Var.	*Success*	Binary (success = 1 if round is filled, vice versa)
Independ. Var.	*Comment Quantity*	How many comments were posted during the campaign
	Comment Sentiment	Degree of comment sentiment (1 = positive, vice versa)
	Comment Length	How many bytes does a comment entail
	Reply Ratio	The number of founder's response vs. comments
	Reply Length	How many bytes does a reply entail
	Reply Speed (hour)	The time lapsed for the founder to respond to a comment
Control Var.	*Fundraising Goal*	The number of funds to be raised for a campaign
	Exper. as Founder	How many campaigns has a founder initiated
	Exper. as Investor	How many campaigns has a founder invested in
	Duration (days)	The length of a campaign
	Video	If the campaign offers a video component
	Picture	How many pictures does the campaign provide
	Update	How many updates founders post during the campaign

Table 6. Binary logistic regression results in a correlation between variables and success [5].

Model	**#1**		**#2**		**#3**	
Variable	**Coeffic.**	**Signific.**	**Coeffic.**	**Signific.**	**Coeffic.**	**Signific.**
Constants	2.357***	.0.000	3.180***	.000	2.676***	.000
Control Variables						
Fundraising Goal (log.)	-0.575***	.000	-0.821***	.000	-0.834***	.000
Duration (log.)	-1.098***	.000	-1.411***	.000	-1.317***	.000
Num of Video (log.)	0.450**	.002	0.440***	.005	0.443***	.004
Num of Picture (log.)	0.013	.958	-0.434*	.093	-0.472*	.070
Update	0.339	.131	0.323	.179	0.362	.132
Experience as Founder	0.297***	.000	0.314***	.000	0.306***	.000
Experience as Investor	-0.282***	.000	0.225***	.001	0.228***	.001
Independent Variables						
Comment Quantity			**0.086*** **	.000	**0.285*** **	.004
Comment Sentiment			0.725	.109	**1.319** **	.078
Comment Length			-.002	.418	-0.003	.282
Reply Ratio			0.011	.860	-0.023	.721
Reply Length			**0.012***	.087	**0.012***	.077
Reply Speed			**0.030***	.088	**0.031***	.077
Moderation						

DOI: 10.54254/2755-2721/5/20230670

Table 6. (continued).

Com. Qual. by Senti.			**0.124*** **	.038
Model Fit Metrics				
Model χ^2	**164.28*** **	**272.78*** **	**277.23*** **	
R^2 (Cox&Snell)	.157	.248	.251	
R^2 (Nagelkerke)	.212	.334	.338	
Correct Classification (%)	70.0	74.9	76.1	

Note: *** $p<0.01$. ** $p<0.05$. * $p<0.10$

To classify comment sentiments, a machine learning approach, namely API from BosonNLP, is adopted. The algorithm will output two metrics, the probability of positive sentiment and negative sentiment as demonstrated in Table 4

Similar to section 2.1, logarithmic regression is used to mitigate the polarity of the data set, and independent variables were added in Models 2 and 3 as shown in Table 6, respectively, to assess their effects on the binary fundraising outcomes.

3. Conclusion

To summarize, this paper examines how elements such as bilateral communication between founders and investors could potentially influence fundraising outcome via a comparative analysis. Both Kim and Wang examined the factors that influence crowdfunding campaigns' outcomes and leveraged sentiment analysis to further evaluate the verbal information provided to the audience. However, the focus and methodology are different: Kim et al. considered the audio input in the form of campaign audios which supplements the textual description. Compared to pure text, audio speech provides an additional dimension – the speech style. The geography of interest was primarily the U.S. On the other hand, Wang studied the interactions between founders and investors beyond campaign materials on a crowdfunding platform in China. Although both studies offered valuable insights, there are a few areas to be addressed in future studies. Crowdfunding dynamics, investors' preferences, and speech styles could be drastically different across geographies. Additionally, communications both inside and outside of the platform should both be included in the model. For example, if a founder and an investor are first connected on a crowdfunding platform, take the communications to offline channels (i.e., Facebook), and share the campaign in a personal network, such a sequence of action should be captured.

Furthermore, an aspect missing in both analyses is the breakdown of the contribution of investors. For instance, for those who have "succeeded" in completing the round, what is the average investment ticket size? Is the round filled by 20 major investors, or it's filled by 200 smaller investors? In the former case, the founder likely should spend more time appealing to and interacting with each investor. Following this logic, the frequency, duration, and quality of their communication should be assigned a heavier weight in the model. In the latter case, how popular the fundraising campaign is on social media platforms – for instance, the number of shares on Facebook and retweets on Twitter – should be assigned more weight. To summarize, there is a wealth of literature that has studied the correlation between fundraising success and various aspects of a campaign, seeking to generate a better prediction. As an extension to existing research, it may be worth implementing technologies such as sentiment analysis to provide guidance in making "persuasive" pitches on a crowdfunding platform. The beauty of LSTM RNNs in text analysis context is to "predict the future" – meaning wording recommendations will be generated. It would also be an effort to democratize access to funding, especially for founders who have solid ideas or strong technical backgrounds but may not be experienced with capital markets. For instance, the wording choice advice in written text and campaign videos, the strategy to best appeal to the target audience, and what key content to highlight.

References

[1] Johan S A and Zhang Y 2020 SSRN Electronic Journal
[2] Wang W, He L, Wu Y J and Goh M 2021 Computers in Human Behavior vol 114

[3] Kickstarter 2022 Retrieved September 29, 2022, from https://www.kickstarter.com/help/stats
[4] Babayoff O and Shehory O 2022 PLOS ONE vol 17
[5] Wang N, Li Q, Liang H, Ye T and Ge S 2018 Electronic Commerce Research and Applications vol 27 pp 106–117
[6] Kim J, Cho D and Lee B 2016 Association for Information Systems
[7] Kunz M M, Bretschneider U, Erler M and Leimeister J M 2016 Electronic Commerce Research vol 17 pp 425–461
[8] Wang W, Guo L and Wu Y J 2022 Technological Forecasting and Social Change vol 174
[9] Mollick E 2014 Journal of Business Venturing vol 29 pp 1–16
[10] Parhankangas A and Renko M 2017 Journal of Business Venturing vol 32 pp 215–236
[11] Mostow J and Beck J 2006 Natural Language Engineering vol 12 pp 195–208
[12] Saura J R, Palos-Sanchez P and Grilo A 2019 vol 11 pp 917