The Mind Behind Crowdfunding: An Empirical Study of Speech Emotion in Fundraising Success

Research-in-Progress

Jongho Kim  
KAIST College of Business  
85 Hoegiro Dongdaemoon-gu Seoul, Korea  
jonghkim@kaist.ac.kr

Daegon Cho  
KAIST College of Business  
85 Hoegiro Dongdaemoon-gu Seoul, Korea  
dgcho@business.kaist.ac.kr

Byungtae Lee  
KAIST College of Business  
85 Hoegiro Dongdaemoon-gu Seoul, Korea  
btle@business.kaist.ac.kr

Abstract

In online crowdfunding, individuals gather information from two primary sources, video pitches and text narratives. However, while the attributes of the attached video may have substantial effects on fundraising, previous literature has largely neglected effects of the video information. Therefore, this study focuses on speech information embedded in videos. Employing the machine learning techniques including speech recognition and linguistic style classifications, we examine the role of speech emotion and speech style in crowdfunding success, compared to that of text narratives. Using Kickstarter dataset in 2016, our preliminary results suggest that speech information—the linguistic styles—is significantly associated with the crowdfunding success, even after controlling for text and other project-specific information. More interestingly, linguistic styles of the speech have a more profound explanatory power than text narratives do. This study contributes to the growing body of crowdfunding research by providing the unexplored aspect of retrieving speech information from the video.

Keywords: Crowdfunding, Social media, Machine learning, Speech recognition
Introduction

“They say video killed the radio star. The question is: Did it also kill the print star? And what does the answer mean for online content marketing?” - Liraz Margalit (quoted in Psychology Today 2015)

In online crowdfunding, individuals gather information from two primary sources, a video pitch and a text narrative as in Figure 1, when the backers make the contributing decision (Frydrych et al. 2014). By the nature of crowdfunding, contribution activities involve risks of the failure of the proposed project. The primary concerns of the risk may arise from the quality of rewards and unsuccessful fundraising of campaigns. In this regard, crowdfunding imposes the risk status where the available information is often restricted or limited. The incomplete circumstance may lead backers to rely more on the emotional aspect when they make a decision (Hanoch 2002; Fedorikhin and Shiv 1999). This context provokes an issue that the information from the video may play a pivotal role in the success of fundraising because videos can facilitate both communication and emotional bonding between the entrepreneur and potential backers (Frydrych et al. 2014; Hekman and Brussee 2013; Mollick 2014; Xu et al. 2013; Moritz and Block, 2016)

Despite the fact that a growing number of crowdfunding projects are presented with the video clip, researchers are more likely to focus on analyzing the text narratives in the context of recommendation systems based on semantic analysis or finding rhetorical techniques (Xu et al. 2014; Pongetti 2011; Beier and Wagner 2015). In contrast, only a few of studies have dealt with video pitches in the crowdfunding context. These studies found that the majority of crowdfunding campaigns – e.g., 86% of projects in Kickstarter – has the video, and the attached video clip significantly increases the success of fundraising (Frydrych et al. 2015; Xiao et al. 2015). However, the prior literature of the video pitch has a clear limitation because they only treated the video pitch as a binary variable, whether or not a project is with the video pitch. However, as one may expect, there are large variations of the video pitch such as view counts, running time, and other idiosyncratic characteristics. Since these important attributes have largely been ignored, the impact of the video characteristics on the fundraising success remains unanswered question. To the best of our knowledge, there has been no previous studies that examine the topic. To supplement this undeveloped research stream and fill the gap in the crowdfunding literature, this study examines the impact of videos in the context of crowdfunding more comprehensively and rigorously than previous studies.

The linguistic styles of speech can affect the perception of backers about the project. For example, in business email communications, people appears to perceive more negative emotions than the positive emotions (Byron 2008). Social propensities that people shows in business communications affect others' perceptions of their occupational proficiency (Barrick and Mount 1991). Also, according to the media richness theory, media formats largely affect the cognitive and emotional level of communication. Therefore, linguistic styles may be an essential part for attracting backers in the crowdfunding context moderated by the different types of media. In this context, more explicitly, our research question is: to what degree do the
speech emotion, speech style, and social propensities perform an additional role to text narratives in crowdfunding success? Our empirical methods borrow the state-of-the-art machine learning techniques such as speech recognition and linguistic style classifications.

Our preliminary results suggest that the linguistic styles of speech and text narratives are good predictors of crowdfunding success by significantly increasing the explanatory power compared to the base model. Still, these preliminary results present the statistical significance of a few variables because of the limited sample size. Our study, however, contributes to the growing body of crowdfunding research in several directions. Firstly, we provide a novel approach to retrieving speech information from the dataset containing video information. To the best of our knowledge, there has been no research in information systems or marketing domains that use the speech recognition technique to extract relevant information from the video which can affect the behavior of market participants. Therefore, this study provides a pioneering approach by suggesting the method to extract relevant information from videos. We believe that our suggested method can further be applied to other studies. Secondly, we introduce a brand new API to extract linguistic styles including the various dimensions of emotions, writing styles, and social propensities which may affect the perception of potential backers to the projects. This advanced text analysis technique helps us to explore the complex nature of the effects of creator generated contents in crowdfunding. Lastly, our results can respond to the unanswered, but still important, question regarding how to design the video pitch and the text narrative, for both crowdfunding creators and platforms.

In the next section, we review related literature. Then, we explain our data and empirical methods that are used for speech recognition, linguistic style extraction, and dimensionality reduction. We present the current preliminary results in the next section. Lastly, we discuss the future direction of our research and managerial implications of this study.

**Literature Review**

**Effects of Media Formats on Contribution Intent**

Our study is closely related to the effects of media format on contribution behavior. Kuppuswamy and Bayus (2015) found that a video can additionally attract a potential donor to support money on a project. Similarly, Beier and Wagner (2015) examined the effects of media formats on the crowdfunding context by borrowing the media richness theory (Daft et al., 1987). They found that the enhanced communication capabilities of videos leads to the higher quality of communication than images and texts. In line with the theory, Liu et al. (2009) found that videos have the highest level of media richness and help to interact best on a cognitive as well as the emotional level of communication. Therefore, prior studies echoed that the video may give the more substantial effect to potential backers than what the other types of media do. However, the previous studies had a clear limitation that the information generated from the video has not been measured rigorously.

A particular project’s success in the crowdfunding context is highly likely to be associated with characteristics of the project provided, even though the factors affecting fundraising may vary in many aspects (Lin et al., 2014). A growing number of studies investigates the key success factors of the project, and they analyzed many possible factors including prior contribution, social exposure, friendship connections, geographic and cultural proximities, soft factors, reciprocity, rhetorical techniques and emotions on the fundraising success (Mollick 2014; Burtch et al. 2013; Li and Duan, 2014; Mollick and Kuppuswamy, 2014).

Among those in the literature, it appears that our study may closely be related to the role of emotions. Capital providers in crowdfunding are not only financially motivated, but social reputation and intrinsic motives can also play a significant role (Lin et al. 2014). Prior studies found that emotional factors are considerably related to the crowdfunding success (Lin et al. 2013). Therefore, with this regard, the video may be one of the most important components, because it can give various aspects of emotions to potential contributors. However, despite the importance of the videos, the effects are rarely examined. In most previous studies, the video was simply treated as a binary variable by which those studies cannot fully consider the richness and the variations of videos, presumably due to the technical challenges and limitations (Kuppuswamy and Bayus 2015; Mollick 2014; Beier and Wagner 2015). In this context, our...
study can fill the important gap in the crowdfunding literature by coping with the various dimensions of the untreated aspects for the video pitch and the text narratives.

Why Does Linguistic Style Matter?

Our study is also associated with the literature that examines how the linguistic features affect funding success in crowdfunding or Peer-to-Peer (P2P) lending context. Larrimore et al. (2011) showed the relationship between language use and persuasion success where unaffiliated individuals borrow money from general public based on text descriptions. They found extended, concrete, and quantitate words are likely to increase funding success. Conversely, informing about the personal details were negatively related with funding success. Gao and Lin (2014) showed that the linguistic styles and positive attitude reveals a higher probability of being funded. Also, they found that certain dimensions of readability and deception cues were related to loan request success. On the other hand, Mitra and Gilbert (2014) found that the prediction accuracy for the project success dramatically increases by accounting language used in text narratives by showing that the error rate dropped from 17.0% to 2.4%. Similarly, Herzenstein et al. (2011) found that being trustworthy or successful in the text are associated with increased loan funding. In short, prior studies commonly found the linguistic features are important factors of being successfully funded. More recently, Gorbatai and Nelson (2015) highlighted the effect of gender and language in the video pitch on the crowdfunding success. Their findings indicated that “female-specific linguistic patterns are preferred over male-specific patterns.” However, despite the fact that the oral speech in the video pitch does have more direct effects than text narratives in terms of linguistic styles, no prior literature have investigated the effects from video information. Therefore, our research contributes to this research stream by examining both the linguistic styles of speech and texts, which can be regarded as different compositions in the linguistic styles.

The Impact of Entrepreneurs’ Pitch on the Success of Project

Delivering an oral presentation, so-called ‘pitch’, has become a common procedure for entrepreneurs who are seeking investment from potential investors. While the pitch is delivered in various circumstances, these presentations are widely used as a method to measure the entrepreneurs’ levels of social competencies and other important human factors (Baron and Brush, 1999; Mason and Harrison 2003). Also, due to the growing importance of the pitch, studies in the entrepreneurship domain measure the impact of the pitch skills and characteristics on investment decisions (Clark 2008). Subsequent studies in this research stream naturally turn their attention to the linguistic patterns and emotional aspects of the pitches (Chen et al., 2009). The video pitch in the crowdfunding context can be regarded as a type of orally presented pitches by which the potential backers may evaluate the creator’s personal characteristics and potential competencies. While the studies in the field of entrepreneurship emphasized the necessity of analyzing the pitch, there has been the lack of research in the crowdfunding, and our study attempts to fill the gap.

Data and Method

Data

We collected a random sample of Kickstarter campaigns drawn from 148,398 Kickstarter campaigns during a two-year period during March 1, 2014 to February 29, 2016. Following the filtering process, we only considered the project that have videos and text description, excluding 55,162 projects and 507 projects for which made it impossible to compare a text and speech information. Also we maintain completed projects, excluding 205 purged projects, 156 projects stopped by Kickstarter due to intellectual property disputes, 17,101 canceled projects, and 1,268 suspended projects. The resulting sample includes 500 crowdfunding campaigns, and the mean of pledged amount of these campaigns is $11,522 from a mean of 169 backers.

This dataset contains a set of variables that can be treated as the key success factors related to characteristics of projects and creators in prior studies as listed in Table 1. We extend these variables by including the
linguistic style and the emotions from the text narratives and the videos. We use the video at the top of the project page as the representative when there are multiple videos.¹

<table>
<thead>
<tr>
<th>Types of Variables</th>
<th>Control Variables</th>
<th>Descriptions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Project Characteristics</td>
<td>fundraising_goal</td>
<td>The minimum amount of funding the project creator want to raise</td>
</tr>
<tr>
<td></td>
<td>desc_length</td>
<td>Number of words in project description</td>
</tr>
<tr>
<td></td>
<td>risk_desc_length</td>
<td>Number of words in project risk section</td>
</tr>
<tr>
<td></td>
<td>funding_duration</td>
<td>Number of fundraising days for the project</td>
</tr>
<tr>
<td></td>
<td>num_photo</td>
<td>Number of photos in the project</td>
</tr>
<tr>
<td></td>
<td>num_video</td>
<td>Number of videos in the project</td>
</tr>
<tr>
<td></td>
<td>length_video</td>
<td>Length of video in seconds</td>
</tr>
<tr>
<td></td>
<td>staff_pick</td>
<td>Whether Kickstarter gives a &quot;project we love&quot; badge</td>
</tr>
<tr>
<td></td>
<td>project_social_expo</td>
<td>Number of external links such as personal website, Facebook, Twitter, etc.</td>
</tr>
<tr>
<td></td>
<td>sure</td>
<td></td>
</tr>
<tr>
<td></td>
<td>num_rewards</td>
<td>Number of rewards in the project</td>
</tr>
<tr>
<td>Creator Characteristics</td>
<td>backing_histroy</td>
<td>Number of projects that the creator has backed</td>
</tr>
<tr>
<td></td>
<td>creating_history</td>
<td>Number of projects that the creator made before</td>
</tr>
<tr>
<td></td>
<td>facebook_connected</td>
<td>Whether the creator shares own Facebook account or not</td>
</tr>
<tr>
<td></td>
<td>gender</td>
<td>Predicted gender of creator based on her/his first name²</td>
</tr>
</tbody>
</table>

**Speech Recognition Process**

Speech is a continuous audio stream which is often described as dynamic states with some degree probabilistic without explicitly distinguished parts, meaning that there are no certain boundaries between words. The traditional procedures to uncover the text from a speech is as follows: 1) split the speech waveform by silences to obtain utterances 2) find the best matching combinations of words for each utterance through mathematical models including an acoustic model, a phonetic dictionary and a language model (Walker et al. 2004). The majority of speech recognition systems such as Dragon, Harpy and Sphinx are based on Hidden Markov Models (HMM) which enable us to treat a speech signal in the matching process (Baker 1975; Lowerre 1976; Walker et al. 2004). However, around 2012, Deep Neural Networks (DNNs) revolutionized the field of speech recognition. Especially, in this paper, we use Long Short-term Memory Recurrent Neural Networks (LSTM RNNs) which has better performance than outperforms DNNs by having additional recurrent connections and memory cells. In this regard, we apply Google Speech API to extract a text from a speech which currently used in Youtube automatic subtitle. (Sak et al. 2014)

Note that there are still technical limitations of the speech recognition. Especially noteworthy is the source separation problem, also known as the cocktail party problem, where a number of sounds are mixed together and the objective is to recover the original component signals. However, the purpose of using a video in the crowdfunding campaign is to inform the clear information to potential backers, which relieve

---

¹ As a preprocessing, we transform the video data into audio files using FFmpeg who provides set of libraries dealing with the multimedia data.

² We use open source library, gender-detector 0.0.4, which deploy the data sourced from Open Gender Tracker's Global Name Data.
concerns about the noise of video. We anticipate that the performance of speech recognition would not be critically harmed by the other noisy sounds. Figure 2 presents sample results of speech recognition.

![Figure 2. A Sample Snapshot of Speech Recognition using Google Speech API](image)

**Extracting Emotion, Writing Styles and Social Propensities**

Creators can show various traits of projects through linguistic styles of the speech or text narratives. Our study borrows the pioneering tool for text mining, IBM Watson Tone Analyzer which is built with extensive researches and very recently introduced in February 2016. As shown in Figure 3, Tone Analyzer enables us to extract emotion, writing styles and social propensities: emotion (anger, disgust, fear, joy or sadness), language style (analytical, tentative or confident) and social propensities (openness, conscientiousness, extraversion, agreeableness or emotional range) (Barrick et al. 1991; Gou et al. 2014; Jian et al. 2014).

![Figure 3. Tone Analyzer interface](image)

---

3 We have tried independent component analysis (ICA), one of the most popular methods to separate music and voice, but the application of ICA harms the accuracy of speech recognition by losing the important features of speech. (Makino et al. 2007)


To deal with the multicollinearity problem, we chose to identify the effects of orthogonal linguistic styles on the crowdfunding success, rather than using a predetermined dimensions of linguistic styles from IBM Watson Tone Analyzer. We perform the principal component factor analysis (PCA) for each linguistic styles of speech and text. We retain the components based on the inspection of the correlation matrix eigenvalues that have at least equal to one as in Kaiser (1960). Our analysis identifies the following four distinct components for speech and five components for text description. Some factors based on the absolute value in a decreasing order are presented as:

1. Speech
   A. **Analytic Speaking**: sadness (0.4299), extraversion (0.4119), analytical (0.3835), agreeableness (0.2509), openness (-0.2191)
   B. **Extrovert Speaking**: agreeableness (0.4326), extraversion (0.3892), conscientiousness (-0.3928)
   C. **Confident Speaking**: confident (0.7852), emotional range (-0.3211)
   D. **Disgust Speaking**: anger (0.4306), disgust (0.4014), joy (-0.5363)

2. Text Narrative
   A. **Analytic Description**: emotional range (0.3241), analytical (0.2111), conscientiousness (-0.272)
   B. **Extrovert Description**: extraversion (0.4045), agreeableness (0.4511), joy (0.396), emotional range (-0.399), tentative (-0.3932)
   C. **Mournful Description**: sadness (0.5423), anger (-0.2455)
   D. **Funny Description**: joy (0.2344), disgust (-0.2933)
   E. **Fearful Description**: sadness (0.8091), disgust (0.7207), anger (0.5455), fear (0.5302)

Note that principal component analysis produced different components for speech and text, presumably speech and text have the significantly different roles in information transmission to the backers. In the next section, we explore the impact of different linguistic styles on fundraising success.

**Preliminary results**

We employ Poisson model to deal with the highly skewed distributions of pledged amounts. Also, we include project characteristic, creator characteristics and dummy variables of project category as control variables. In Table 2, we begin by estimating our parameters progressively adding the variables of interest from a model with control variables only. In Model 2, we add the linguistic styles of text narratives. The difference in explanatory power between the base model indicates the text narratives are significant predictors of total pledged amounts.(Model 1 vs. Model 2: ΔPseudo $R^2=0.032$) The result shows the significant negative impact of fearful descriptions which have factors of higher fear and lower funny. Not surprisingly as predicted by media richness theory, in model 3, the linguistic style of speech has more profound explanatory power than the text narratives. (Model 2 vs. Model 3: $Δ$Pseudo $R^2=0.040$) The result presents the positive impact of analytic speaking which has the factors of higher analytic and lower openness. Even though, these preliminary results only show statistical significance of a few variables because of our limited sample size, our results suggest that the linguistic styles of speech and text narratives are powerful predictors of crowdfunding success by significantly increasing explanatory power compared to the base model.

**Table 2 Impact of speech and text on fundraising success**

---

This study is a research about the effect of money harms the crowdfunding success. Additionally, our empirical approaches are valuable to answer on very realistic and practical questions for both entrepreneurs and platform.

<table>
<thead>
<tr>
<th>Category</th>
<th>VARIABLES</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speech</td>
<td>Analytic Speaking</td>
<td>0.202**</td>
<td>(2.94)</td>
<td>0.213***</td>
<td>(3.56)</td>
</tr>
<tr>
<td></td>
<td>Extrovert Speaking</td>
<td>0.091</td>
<td>(1.80)</td>
<td>0.173</td>
<td>(1.92)</td>
</tr>
<tr>
<td></td>
<td>Confident Speaking</td>
<td>0.143**</td>
<td>(2.80)</td>
<td>0.099</td>
<td>(1.51)</td>
</tr>
<tr>
<td></td>
<td>Empathic Speaking</td>
<td>-0.122</td>
<td>(-1.02)</td>
<td>-0.197</td>
<td>(-1.86)</td>
</tr>
<tr>
<td>Text Narratives</td>
<td>Analytic Description</td>
<td>-0.015</td>
<td>(-0.21)</td>
<td>-0.116</td>
<td>(-0.24)</td>
</tr>
<tr>
<td></td>
<td>Extrovert Description</td>
<td>-0.163</td>
<td>(-1.60)</td>
<td>-0.165</td>
<td>(-1.72)</td>
</tr>
<tr>
<td></td>
<td>Mournful Description</td>
<td>-0.022</td>
<td>(-1.10)</td>
<td>0.012</td>
<td>(0.15)</td>
</tr>
<tr>
<td></td>
<td>Funny Description</td>
<td>0.148</td>
<td>(1.02)</td>
<td>0.131</td>
<td>(1.16)</td>
</tr>
<tr>
<td></td>
<td>Fearful Description</td>
<td>-0.390**</td>
<td>(-2.60)</td>
<td>-0.374**</td>
<td>(-2.89)</td>
</tr>
<tr>
<td>Characteristics</td>
<td>Project Characteristics</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>Creator Characteristics</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>Category Control Variables</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>Observations</td>
<td>500</td>
<td>500</td>
<td>500</td>
<td>500</td>
</tr>
<tr>
<td></td>
<td>Pseudo R2</td>
<td>0.401</td>
<td>0.433</td>
<td>0.473</td>
<td>0.520</td>
</tr>
</tbody>
</table>

**Note:** t statistics in parentheses

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

**Discussion and Future Work**

Crowdfunding has become one of the most important online platforms by providing channels to raise money from the general public. In this study, using the speech recognition technique, we attempted to investigate to what degree does the linguistic styles of speech play a role in the crowdfunding success compared to text narratives. Our principal component analysis shows that the speech has a different composition of linguistic styles compared to the text narratives, speech and text may play different roles. We also found that both speech and text narratives are significant predictors of crowdfunding success. In addition, the results show that analytic speaking leads to the positive performance while fearful description harms the crowdfunding success.

This study is a research-in-progress. While we strengthen our empirical work, we will address additional questions by exploring further detail: 1) Which properties of speech will make the project more viral? While the vast of literature seek to find the factors that enable viral marketing, speech information wasn’t examined extensively. To answer this question, we will further extend our dataset into the random sample of the 4,500 projects including the number of shares on Facebook, Twitter and other social media. Additionally, it is valuable to consider 2) For how do the importance of linguistic features of speech and text narratives vary according to the demographic of creators? Even though, Crowdfunding is known as democratic path of collecting the capital, prior literatures have found that the racial disparities or gender do affect fundraising success. (Greenberg and Mollick, 2014; Gorbatai and Nelson, 2015) Therefore, our examination of the interplay between linguistic features and demographics would strengthen any findings about the effect of gender or race on fundraising success.

Our empirical approach not only has the potential to contribute to the growing body of crowdfunding studies, but it can also be used in other research areas. For instance, electronic commerce websites are increasingly adopting online videos to advertise their products. But there is little empirical guidance on how to speak in video or how to design text descriptions. In that perspective, our empirical approaches are valuable to answer on very realistic and practical questions for both entrepreneurs and platform.
References


The Mind Behind Crowdfunding


Andonized Field Experiment.”


