Do User Entrepreneurs Speak Different? Applying Natural Language Processing to Crowdfunding Videos

Jermain Kaminski
RWTH Aachen University
Technology Entrepreneurship Group
TIME Research Area
kaminski@time.rwth-aachen.de

Frank Piller
RWTH Aachen University
Technology and Innovation Management
TIME Research Area
piller@time.rwth-aachen.de

Yibin Jiang
RWTH Aachen University
Technology Entrepreneurship Group
TIME Research Area
yibin.jiang@rwth-aachen.de

Christian Hopp
RWTH Aachen University
Technology Entrepreneurship Group
TIME Research Area
hopp@time.rwth-aachen.de

CHI'17 Extended Abstracts, May 6-11, 2017, Denver, CO, USA.
ACM 978-1-4503-4656-6/17/05.
http://dx.doi.org/10.1145/3027063.3053223

Abstract
UPDATED—February 19, 2017. In this work, we analyze video pitches of creators in crowdfunding campaigns and focus on linguistic particularities of 'lead user' entrepreneurs. Theory suggests that lead users sense needs long before they become known to the broader public and would benefit greatly from finding a solution to these needs. For our study, we consider 404 video pitches of creators on Kickstarter, distinguishing lead users and regular campaigners. The study employs natural language processing (NLP) of (video-to-)speech-to-text content. Initial results indicate that lead users are more oriented towards product and problem-solving rather than focusing on pecuniary motives.

Author Keywords
Crowdfunding; natural language processing (NLP); video; lead users

ACM Classification Keywords
H.5.2 [User Interfaces]: Natural language; H.5.3 [Group and Organization Interfaces]: Web-based interaction.

Introduction
The angel investor Ron Conway (SV Angel) said in open online education class `How To Start A Startup -`

Kickstarter Projects
537 Kickstarter campaigns with expert-evaluation of lead-userness.
↓
Campaign Videos
404 valid campaigns with speech content in videos.
↓
Video to Speech
Conversion of video to audio with `ffmpeg`, if no subtitles (*.vtt) is provided.
↓
Speech to Text
Conversion of video to audio with `Google Cloud Speech API`, as of November 2016 (Python integration).
↓
Text Processing
Lowercase conversion, stopword cleaning, frequency threshold.
↓
NLP Applications
n-grams, `LIWC` classification.

`CS183B` at Stanford University:[1]: *"Literally while you are talking to me in the first minute I am saying: Is this person a leader? Is this person rightful, focused, and obsessed by the product? I am hoping, because usually the first question I ask is: What inspired you to create this product? - I'm hoping that it's based on a personal problem that that founder had and this product is the solution to that personal problem."*

Conway emphasizes the vital, but often neglected role of personal founder aspirations embedded in communication patterns. These affect the perception of others and evidently also the chances for future funding. Our research therefore aims to analyze two under-researched areas:

1. Is language in videos predictive for funding outcomes in crowdfunding, and
2. to what extent are creators on the path to solve a personal problem (lead users) more successful in crowdfunding.

Until January 2017, 118,000 campaigners on the crowdfunding platform Kickstarter [8] successfully raised more than $2.47bn. The platform emerged as one of the gateways for entrepreneurs to present and fund novel and innovative ideas. Through its transparency and open information flow, crowdfunding platforms provide researchers with the opportunity to study recipes of success in these environments. Firstly, the creator's communication and entrepreneurial journey is very transparent, and secondly, crowdfunding investors show a high level of professionalism in making investment decisions [12].

[1] http://startupclass.samaltman.com/, Lecture 9 - How to Raise Money (Marc Andreessen, Ron Conway, Parker Conrad), Timestamp: `1:15`.

Particularly interesting in the crowdfunding setting are lead users. Lead users "anticipate relatively high benefits from obtaining a solution to their needs" and "are at the leading edge of an important market trend" [19]. Lead users are the source of many successful new products [18, 20], for instance in health, technology and sports (cf. Fig. 1 and 2 for examples of 'lead-userness'). They turn into 'user entrepreneurs' [16], when they realize that their inventions have an appeal for other users [6, 2]. The lead user phenomenon appears to be at odds with the dominant paradigm and logic in innovation management and policy. Here, the common perception is the that there are manufacturers who innovate, by identifying an open need in a market, and by engaging in creativity and problem solving techniques, they translate this need into a solution. Not surprisingly then, there is currently little research that can help to shed more light on the lead user phenomenon and its societal importance. We think that crowdfunding platforms are an ideal place to explore lead-userness and how far funding outcomes are driven by the language and content they use to communicate.

An increasing corpus of the crowdfunding literature analyzes project descriptions of crowdfunding campaigns [10, 5, 1, 3]. For instance, Mitra and Gilbert (2014) analyze Kickstarter project descriptions find that the language used by creators to pitch their ideas (phrases) account for 58.56% of the variance around success.

While research indicates that the presence of a pitch video in a Kickstarter project largely influences a campaign's outcome [10, 11], only little work has been done on crowdfunding videos. As for instance an analysis of emotions *to* campaign videos [15]. Entrepreneurship researchers already studied decision making processes of business angels and the preparedness of founders by analyzing television shows

Figure 1: Arpeggio: The Portable Arpeggiator, Sequencer, and Synth: "Ever since I've been fascinated with musical instruments. I get frustrated with standard step sequencers because they aren't capable of executing many of the kinds of melodies I find compelling". Funded, $94,713. Image credit: Tangible Instruments.

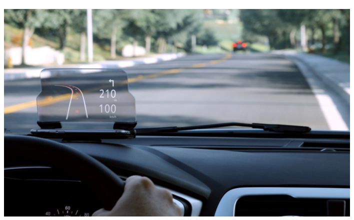

Figure 2: HUDWAY Glass: keeps your eyes on the road while driving: "The idea came from Alex's professional hobby - rally racing. In rally, they have co-pilots for efficient navigation. But co-pilots are humans who can naturally make mistakes". Funded, $622,785. Image credit: HUDWAY Glass.

[9, 14]. However, video pitches by founders in crowdfunding represent a different setting, where creators address "product-investors", i.e. users, for support, as compared to "company-investors" in venture capital (VC).

In this early-stage work, we analyze video pitches of creators in crowdfunding, and focus on linguistic differences of lead user entrepreneurs. By doing so, we extend the research on conversation and emotion in crowdfunding campaigns, moving from written text [10] to speech. As the presentation in a video allows campaign creators to pitch their own motivation for creating the product, they may engage in storytelling to market their product, using lead user terminology.

Data

Our sample consists of 537 "Product Design" or "Technology" Kickstarter projects in the United States, initiated between September 1st and October 31st 2015. Using Python library 'Beautiful Soup', we automatically downloaded Kickstarter campaign videos that are uploaded by creators to substantiate their descriptions. We receive a total of 404 valid videos, as some projects have either no video present or no spoken word (music only).

Video to Text

In our research, all related mp4 videos are firstly converted into audio format. The average video length in sample is 2:46 min. We used ffmpeg to convert videos in mp4 format into audios in wav format with a bit rate of 160 kbit/s, and a sampling rate of 44100 Hz with 2 channels. We use a Python implementation of the `Google Speech API`, `SpeechRecognition`, to further process the audio files and convert them into text. Each audio is split into snippets of each 10 seconds and then stored as plain text. We converted all video-to-text files into lowercase and tokenized them to unigram, bigram and trigrams. We also removed stop words and numbers, and only consider phrases with a frequency of equal of greater than 5.

Lead Users

In order to evaluate lead-userness, we recruited a total of 27 individuals to assess the campaign and founder characteristics of the selected videos through a survey comprising multiple questions on campaigners, campaigns, and product characteristics. Applying several robustness checks, the intraclass correlation coefficient varies between 0.89 and 0.96. Following Franke, von Hippel, and Schreier (2006),[4] we asked raters, whether "The creator has ideas that are 'ahead of the trend'". Similarly, we assessed the 'benefit and utility expected' from the product. We further assessed whether campaign founders are dissatisfied with existing product solutions on the market and whether or not they believed that there exists an unresolved problem. We conclude 'lead-userness', if the average score of weighted items on 'benefit expected' and 'ahead of trend' succeeded or equals the value 5 on a 7-p Likert scale, with 7 = 'strongly agree' and 1 = 'strongly disagree'. Cronbach's alpha is 0.79. Based on this value, we divided our total sample into two cohorts, that are (1) lead users (LU), whose value ≥ 5, and non-lead user (NLU) whose value < 5. As shown in Table 1, we get 102 lead users and 302 non-lead user, each differentiated by funding outcome. Our final sample comprises 604 unique phrases for 102 LU and 1707 unique phrases for 302 NLU. While responders rated lead-userness based on all available campaign information, there is potential for mimetic behavior and false-positives, i.e. creators pretending to have lead user motivations.

Method

Our analysis follows the approach of Mitra and Gilbert (2014) [10], who use textual descriptions of Kickstarter

	LU	NLU
Success	56	131
Failed	46	171
Total	102	302

Table 1: Data samples.

	LU β
Duration	1.92
Comments	0.09
Total_Words	-0.47
Backers	-0.58
Goal	-0.60
	NLU β
Duration	2.01
Comments	-0.12
Backers	-0.41
Total_Words	-0.49
Goal	-0.52

Table 2: Control variables with non-zero predictive power for funding outcome (Z-scores).

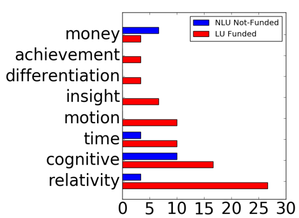

Figure 3: Selected LIWC categories. The x-axis depicts the % to which phrases predict LU or NLU projects.

projects to extract phrases that predict funding outcomes on Kickstarter. In our analysis, we also apply a penalized logistic regression that accounts for collinearity and sparse features. Using the `Python` library `sklearn`, we predict the dependent variable *funded*, implying 'status: *successful*', i.e. *\$ amount pledged* $\geq$ *\$ fund-raising goal* within a set time frame. The parameter α, determining the movement of the weights of the predictive variable, is set to α = 1.0. As such, our regression represents a Lasso model (as compared to α = 0.0, a Ridge regression model).

Results

Table 3 and 4 present the parameter β, as generated by our model. We list the top 30 phrases with positive β weights and top 30 phrases with negative β weights for LU, and NLU respectively. As for lead users (LU), Table 3 lists the top 30 phrases having positive β weights that predict a positive funding outcome and top 30 phrases having a negative β weight.

Table 3 provides an indication for *roughly* four dimensions that are represented in successful LU projects:

1. **Product**: "batteries" (β 7.06), "bag" (β 6.43), "camera" (β 2.40).
2. **Utility**: "accessible" (β 6.67)), "affordable" (β 3.48), "bigger" (β 3.25).
3. **Problem solving**: "allows" (β 3.56), "complicated, "add" (β 3.26), (β 2.69), "can't" (β 1.77).
4. **Experimental**: "custom" (β 1.80) , "explore" (β 1.56).

Terms in our sample further indicate that LU products address certain user groups specficially, such as chefs (β 5.42), aged people (β 3.51), dogs (β 2.23), babies ("diaper" β 2.20), or open source technology users, such as "android" (β 2.55). As compared to non-successful LU and NLU (Table 4), LU build their product upon "already" (β 2.24) existing solutions ("based" β 1.72) or "add" (β 3.26) functionality to something that is known to the user. All in all, inventions based on lead user component result in rather incremental (competence-enhancing) solutions [17, 7]. In Table 2, we show control variables with non-zero predictive power in funding outcomes.

To test for the robustness of our results, we also applied a Linguistic Inquiry and Word Count (LIWC) [13] analysis [10]. `LIWC` uses a hand-built dictionary to conduct text analysis on given input text. The library analyzes texts and calculates a percentage value that measures to what extent words in a selected dataset represent semantic categories. We identified a total of 38 LIWC categories and limit our report to eight that showed significant differences among successful LU and non-successful NLU campaigns (Figure 3). According to LIWC, LU differentiate most on positive loadings for words related to "relativity", "motion" and "time", as well as within the "cognitive processes" category. Also, LU seem to exhibit less focus on pecuniary motives and report more on prior achievements, differentiation points, as well as insights into their product. This corroborates prior results depicted in Table 3 and 4. According to our current analysis, they show no differences in "Affective Processes", i.e. positive or negative emotions.

Conclusions and Future Work

Considering that campaign creators engage in various ways to emotionally appeal to potential investors, it might be important to extend our results along the lines of impression management and emotional assuage. Ongoing research is therefore focused on a more nuanced analysis of speech-

(LU) phrases	β	*(LU) phrases*	β
batteries	7.06	block	-5.73
accessible	6.67	apple	-3.80
bag	6.43	almost	-3.39
chef	5.42	balance	-3.35
behind	4.17	attached	-3.20
build	4.12	across	-2.97
allows	3.56	ability	-2.47
age	3.51	capture	-2.33
affordable	3.48	application	-2.33
add	3.26	company	-2.32
bigger	3.25	access	-2.12
complicated	2.69	colors	-2.08
addition	2.66	built-in	-2.03
code	2.62	achieve	-1.97
android	2.55	another	-1.76
comes	2.50	engineering	-1.74
camera	2.40	babies	-1.74
brought	2.34	amazing	-1.73
creator	2.28	diapers	-1.70
already	2.24	bottom	-1.69
dog	2.23	always	-1.61
diaper	2.20	believe	-1.60
along	1.99	inch	-1.45
ago	1.81	fit	-1.40
custom	1.80	box	-1.35
best	1.80	away	-1.28
can't	1.77	extra	-1.28
based	1.72	battery	-1.26
half	1.68	measure	-1.23
explore	1.56	edge	-1.19

Table 3: Left: Top 30 phrases predicting positive outcome of LU crowdfunding projects. Right: Top 30 phrases predicting negative outcome of LU crowdfunding projects. Z-score normalized.

(NLU) phrases	β	*(NLU) phrases*	β
alternative	9.84	bill	-5.48
audio	8.30	bikini	-5.37
aware	6.16	chili	-5.15
click	5.35	cartridge	-3.85
inflatable	5.10	bought	-3.80
add	5.04	dirt	-3.71
arm	4.64	cookie	-3.42
ability	4.64	base	-3.39
app	4.56	bands	-3.30
alone	4.48	calendar	-3.14
books	4.41	blocks	-3.05
allowing	4.39	actual	-3.01
action	4.26	basically	-2.98
anything	4.20	book	-2.97
affordable	3.73	child	-2.91
bikes	3.67	controllers	-2.85
cutting	3.57	characters	-2.82
artist	3.56	boys	-2.82
helmet	3.49	android	-2.81
big	3.30	covers	-2.66
busy	3.23	balance	-2.65
bike	3.22	comfortable	-2.64
begin	3.11	current	-2.64
changing	3.00	attached	-2.59
animals	2.99	batteries	-2.56
bella	2.98	allowed	-2.49
coffee	2.96	girls	-2.49
addition	2.91	attach	-2.46
heat	2.71	hoping	-2.44
fresh	2.70	mounting	-2.39

Table 4: Left: Top 30 phrases predicting positive outcome of NLU crowdfunding projects. Right: Top 30 phrases predicting negative outcome of NLU crowdfunding projects. Z-score normalized.

to-text content (n-grams, t-SNE word embeddings), as well as objects and emotions in videos. Not only are we further interested in mapping out whether campaign representations are mirrored by campaign founders skills, but also we aim at testing whether they can live up to their promise of bringing the product to market, i.e. are able to commercialize their project. Similarly, one may ask whether crowdfunders find the product as appealing as initially suggested. Sentiment analyses of forum comments on the received product, for example, could illumine our understanding further.

Overall, our initial results encourage to continue the research on user entrepreneurs in crowdfunding. Two potential tools resulting of such research could be (1) an audio and video input powered prediction engine for start-up pitches, in particular in the crowdfunding setting, and (2) an algorithm highlighting crowdfunding projects with apparent lead user motivation.

Acknowledgements

We would like to thank our reviewers, as well as colleagues at the MIT Media Lab and the TIME Research Group at RWTH Aachen for several comments that improved our work. Last but not least, we like to thank Oleksandr Zastupailo for preparing the data and all 27 experts for their precious contribution to this project.

References

[1] Tim Althoff, Cristian Danescu-Niculescu-Mizil, and Dan Jurafsky. 2014. How to ask for a favor: A case study on the success of altruistic requests. *arXiv preprint arXiv:1405.3282* (2014).

[2] Erkko Autio, Linus Dahlander, and Lars Frederiksen. 2013. Information exposure, opportunity evaluation, and entrepreneurial action: An investigation of an online user community. *Academy of Management Journal* 56, 5 (2013), 1348–1371.

[3] Nihit Desai, Raghav Gupta, and Karen Truong. 2016. Plead or Pitch? The Role of Language in Kickstarter Project Success. (2016). http://nlp.stanford.edu/courses/cs224n/2015/reports/15.pdf.

[4] Nikolaus Franke, Eric Von Hippel, and Martin Schreier. 2006. Finding commercially attractive user innovations: A test of lead-user theory. *Journal of Product Innovation Management* 23, 4 (2006), 301–315.

[5] Michael D Greenberg, Bryan Pardo, Karthic Hariharan, and Elizabeth Gerber. 2013. Crowdfunding support tools: predicting success & failure. In *CHI'13 Extended Abstracts on Human Factors in Computing Systems*. ACM, 1815–1820.

[6] Stefan Haefliger, Peter Jäger, and Georg Von Krogh. 2010. Under the radar: Industry entry by user entrepreneurs. *Research Policy* 39, 9 (2010), 1198–1213.

[7] Rebecca Henderson. 1993. Underinvestment and incompetence as responses to radical innovation: Evidence from the photolithographic alignment equipment industry. *The RAND Journal of Economics* (1993), 248–270.

[8] Kickstarter.com. 2017. Statistics. (2017). https://www.kickstarter.com/help/stats.

[9] Andrew L Maxwell, Scott A Jeffrey, and Moren Lévesque. 2011. Business angel early stage decision making. *Journal of Business Venturing* 26, 2 (2011), 212–225.

[10] Tanushree Mitra and Eric Gilbert. 2014. The language that gets people to give: Phrases that predict success on kickstarter. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*. ACM, 49–61.

[11] Ethan Mollick. 2014. The dynamics of crowdfunding: An exploratory study. *Journal of Business Venturing* 29, 1 (2014), 1–16.
[12] Ethan Mollick and Ramana Nanda. 2015. Wisdom or madness? Comparing crowds with expert evaluation in funding the arts. *Management Science* 62, 6 (2015), 1533–1553.
[13] J. W. Pennebaker and M. E. Francis. 1999. Linguistic Inquiry and Word Count. Lawrence Erlbaum.
[14] Jeffrey M Pollack, Matthew W Rutherford, and Brian G Nagy. 2012. Preparedness and cognitive legitimacy as antecedents of new venture funding in televised business pitches. *Entrepreneurship Theory and Practice* 36, 5 (2012), 915–939.
[15] Josue Reyes and Cristina Bahm. 2016. Crowdfunding: Applying Collective Indexing of Emotions to Campaign Videos. In *Proceedings of the 19th ACM Conference on Computer Supported Cooperative Work and Social Computing Companion*. ACM, 385–388.
[16] Sonali K Shah and Mary Tripsas. 2007. The accidental entrepreneur: The emergent and collective process of user entrepreneurship. *Strategic Entrepreneurship Journal* 1, 1-2 (2007), 123–140.
[17] Michael L Tushman and Philip Anderson. 1986. Technological discontinuities and organizational environments. *Administrative science quarterly* (1986), 439–465.
[18] Glen L Urban and Eric Von Hippel. 1988. Lead user analyses for the development of new industrial products. *Management Science* 34, 5 (1988), 569–582.
[19] Eric Von Hippel. 2005. Democratizing innovation: The evolving phenomenon of user innovation. *Journal für Betriebswirtschaft* 55, 1 (2005), 63–78.
[20] Eric Von Hippel, Susumu Ogawa, and Jeroen PJ De Jong. 2011. The age of the consumer-innovator. *MIT Sloan Management Review* 53, 1 (2011), 27.