Roles of Text in Images in Crowdfunding Projects

Liu Yue (A0092394X)

1. Project Objectives Description

Crowdfunding is a relatively new form of informal venture financing. By an open call for supporters through the Internet, entrepreneurs are able to solicit financial supports from a large number of distributed participants for various causes ranging from for-profit, social or cultural (Belleflamme, Lambert & Schwienbacher, 2014). Currently, there exist more than 2,000 online platforms that facilitate crowdfunding transactions.

Crowdfunding models can be categorized as either donation-based, reward-based, equity-based, or lending-based depending on the nature of returns provided to supporters (Lam & Law, 2015). For equity-based and lending-based crowdfunding such as Sellaband and Prosper, supporters receive financial return in terms of shares or interests (Agrawal, Catalini & Goldfarb, 2015; Zhang & Liu, 2012). For donation-based crowdfunding such as JustGiving, supporters receive no tangible returns (Smith, Windmeijer & Wright, 2015). For reward-based crowdfunding such as Kickstarter, supporters receive non-financial rewards described in the crowdfunding campaigns. The rewards are usually in the form of a token gift of appreciation for backing a project or pre-ordering a product or service (Kuppuswamy & Bayus, 2015).

In 2014, crowdfunding platforms raised $16.2B which is an increase over the $6.1B raised in 2013 (Kuppuswamy & Bayus, 2015). Kickstarter, which is the leading crowdfunding platform in the US, has raised over $2.6B in pledges from 12M backers to fund almost 114,000 creative ideas (Kickstarter, 2016). Kickstarter implements an 'All-or-Nothing' (AON) funding rule, where the entrepreneurs set a fundraising goal and keep no funding unless the goal is successfully reached (Schniter, Sheremeta & Shields, 2015). Hence, understanding the drivers of fundraising success is important for entrepreneurs to formulate crowdfunding campaign.

To date, there have been many empirical studies in this domain focusing on identifying the project features associated with successful funding outcomes. For instance, research shows that crowdfunding success is positively related to narratives in the project description (Mitra & Gilbert, 2014), entrepreneurs' descriptions (Marom & Sade, 2013), reward scheme and goal setting strategies (Xiao, et al., 2014; Li & Jarvenpaa, 2015). Research also shows that crowdfunding success is positively related to entrepreneurs' features such as geography (Mollick 2014), prior experiences (Zvilichovsky, et al., 2013), and personality (Thies, Wessel, Rudolph & Benlian, 2016).

In the vast majority of existing crowdfunding studies, the features studied are mainly engineered from the text in project presentations. Researchers generally do not consider features from the content formats of higher media richness, such as image and video, in project presentations. While some studies attempt to study the association between video duration and video transcript with successful funding outcomes, the information from visual images have not been investigated for their impacts on project success (Beier & Wagner, 2015; Thies, Wessel, Rudolph & Benlian, 2016).

Our interest in this project is to complement the existing research by explicitly considering the roles of text in images in the project presentations of crowdfunding campaigns. The focuses are to investigate the reasons for project creators to use images and to embed text in image, how the narrative of the text in image is different from the text description, and how text in image is associated with project funding success. We believe these focuses worth to be investigated as higher media richness is related to better understanding of information and higher viewer interests (Sun & Cheng, 2007). Since there exists no prior work on text in image analysis, we Instead, we expect that our empirical findings will be useful for future theory-building.

2. Literature Review on Text in Image Extraction

OCR research is comparatively old in the field of pattern recognition research. OCR has been used by libraries and law firms to digitize books and documents for at least 30 years (Mori, Suen & Yamamoto, 1992). Recently, it has been combined with text detection algorithms to extract text from natural scenes, house numbers and business cards (Epshtein, Ofek & Wexler, 2010).

While existing commercial products, such as ABBYY, are in widespread, open source OCR engins such as OCRopus and Tesseract are available and have shone brightly with their results (Breuel, 2008; Smith, 2007). With new pattern algorithms developed to specifically addressing different issues in OCR, the text extraction accuracy has been improved largely. For instance, such algorithms include connected components analysis for identifying distinct letters (Zhong, Karu & Jain, 1995), otsu thresholding for determining word spacing (Abak, Baris, & Sankur,1997), stroke width transform algorithm for detecting regions of text in a language agnostic manner (Epshtein, Ofek, & Wexler, 2010). In addition, different image processing pipelines are designed for text extraction under different scenarios (Sas & Zolnierek, 2013).

3. Research Methodologies

3.1 Context and Data

The empirical setting for our study is Kickstarter, one of the oldest and largest reward-based crowdfunding communities on the Internet. We collect project information from kickstarter.com over the period from April 2009 to February 2016. The dataset includes projects completed in all categories. The dataset consists of 99,538 successful projects and 144,681 failed projects. For each project, an exhaustive list of project-level information that the backers can observe on Kickstarter platform is captured, including project goal, project duration, project deadline, project category, text in project descriptions, and images in project descriptions.

3.2 Methods

3.2.1 Text Extraction from Image

For each project description, text in images is extracted with Optical Character Recognition techniques, using an implementation of the stroke width transform algorithm created by Kwok (2014).

Stroke width transform algorithms was first created by Microsoft research staff Epshtein, Ofek and Wexler, demonstrating the key idea that the 'one feature that separates text from other elements of a scene is its nearly constant stroke width' (2010). In the evaluation for images depicting natural scenes, the algorithm is able to detecting regions of text with precision of 0.73 and recall of 0.60.

The implementation developed by Kowk is a Google Chrome extension called Project Naptha. This JavaScript extension allows text within images to become accessible as text (Kwok, 2014). It accomplishes this by running the algorithm on the images of a Webpage to detect text regions, and then runs those regions through online OCR engines including the GNU Ocrad Software and Terreract (Cocciolo, 2015).

Figure 1.a and 1.b shows this process used on an image of a Kickstarter project, Pebble Smart Watch, with the white boxes identifying the text regions from the images, pink and red boxed identifying words and letters.

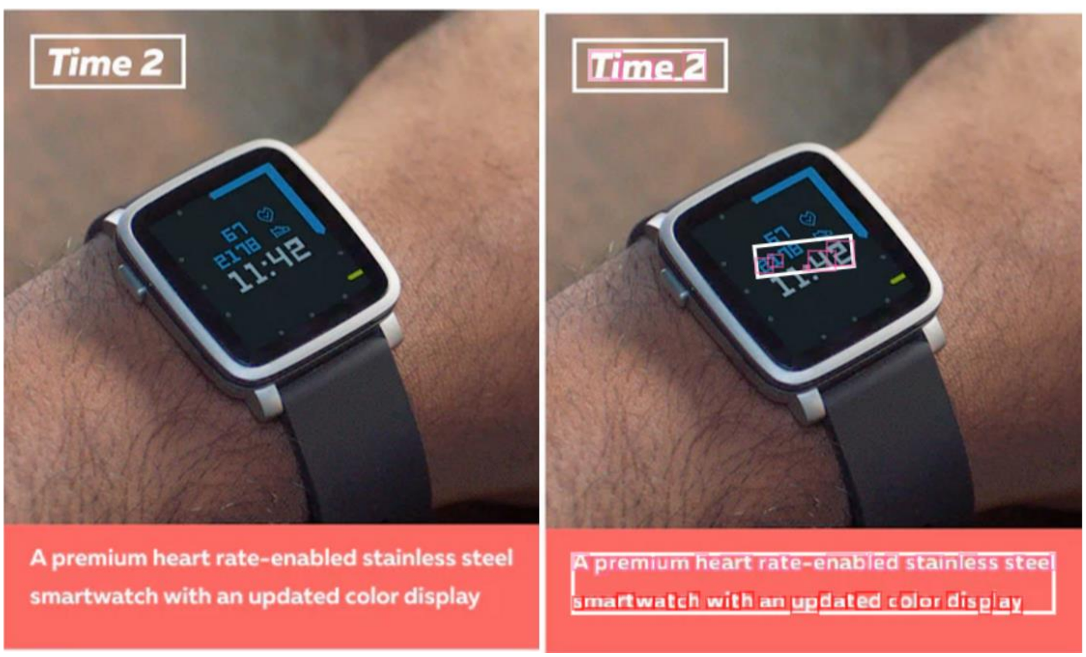

Figure 1: a) Original image; b) Image with text regions and characters recognized in boxes.

The algorithm has minor inaccuracies. For instance, text extracted form Figure 1 is “lrme A premium heart rate-enabled stainless steel smartwatch with an updated color display”. The text box at top left corner should be “Time 2”, but recognized as “lrme”. The text in boxes on the watch face is not recognized as characters at all.

In our project, we plan to wrap the JavaScript extension and use Terresact Engine on a local server. For each Kickstarter campaign, the text from the lists images in project presentation are then extracted and stored.

4.2.2 Spell Correction of Extracted Text

As the text extraction algorithm has may produce inaccurate text output, we conduct spell correction on the text extracted to improve the accuracy and meaningfulness of words extracted.

Before the extracted text can be processed by a spell corrector, we train a spell corrector that is dedicated for OCR outputs that used stroke width transform algorithm for text region detection.

The training process of the spell corrector involves comparing the actual text and OCR output form a training image dataset. We construct training images using common words, with each image differ in styles such as fonts, upper or lower cases, and character spacing. The generation process is similar to the generation of the Adobe Visual Font Recognition dataset (Wang, et al., 2015). The main difference is words are the labeled classes in our dataset instead of font as the labeled classes. Figure 2 is an example of images in the dataset.

Figure 2: Sample images in our training image dataset, with labels of “RESTRAUNT” and “Electrical” respectively.

For each commonly used words, the probability of every possible OCR output text is computed and recorded. For each OCR output character, the probability of the character in the same position in original word, the probability of the former characters and the latter characters in original word are computed.

Using the spell corrector trained, text extracted from Kickstarter campaign images is processed to form meaningful words via algorithm of ngram statistics on words and characters.

4.2.3 Image Classification Based on Functions

Images may have the functions as headings and subheadings, project goal description, product description, reward scheme description, timeline description, entrepreneurs' description and others. We classify them based on the length of text in the image, aspect ratio of the image, and text percentage in the image (i.e. ratio between the area of text boxes identified and area of image).

4.2.4 Differences among Narratives by Phrase Frequencies

For each class of images, a corpus of unigram, bigram and trigram phrases of the words extracted is construed. A large corpus is also constructed by combining the image-specific corpus with a corpus formed with textual description of the project. For each of the corpus, we include only the more common phrases by removing phrases of low frequencies (n>50). This allows us to keep number of observations more than number of independent variables (i.e. phrases), and to have larger number degree of freedom in the model used for funding success prediction.

Alternatively, we can identify top 100 phrases from each corpus, and use them in the model for funding success prediction.

4.2.5 Project Success Prediction

Independent variables include number of images in each image class, length of text in image, and the phrases selected from the corpus. Dependent variable is whether a project is successfully funded. Control variables include campaign features such as project goal, project duration, number reward schemes levels, number of stretch goal levels, whether the project is featured by Kickstarter.

We employ penalized logistic regression to predict whether a project is successfully funded, similar to the research conducted by Mitra and Gilbert in their instigations of funding success predictive phrases (2014). The method is adopted for its well know property of reducing the impact of collinearity and sparsity in the dataset, which are prevalent in phrase dataset (Tibshirani, 1996). In penalized logistic regression, coefficients (i.e. weights of independent variables in the prediction model) are very large on the most predictive features and very small on the less predictive features.

While in Mitra and Gilbert's research, Lasso regression is selected to save computational resources, we can alternatively vary the penalization degrees in the same elastic net family.

5. Preliminary Explorations

The number of images present in the campaign presentation is positively correlated with funding success, especially in more recent years. This can be shown in Figure 3, which indicates successfully funded projects have more images in their campaign presentations.

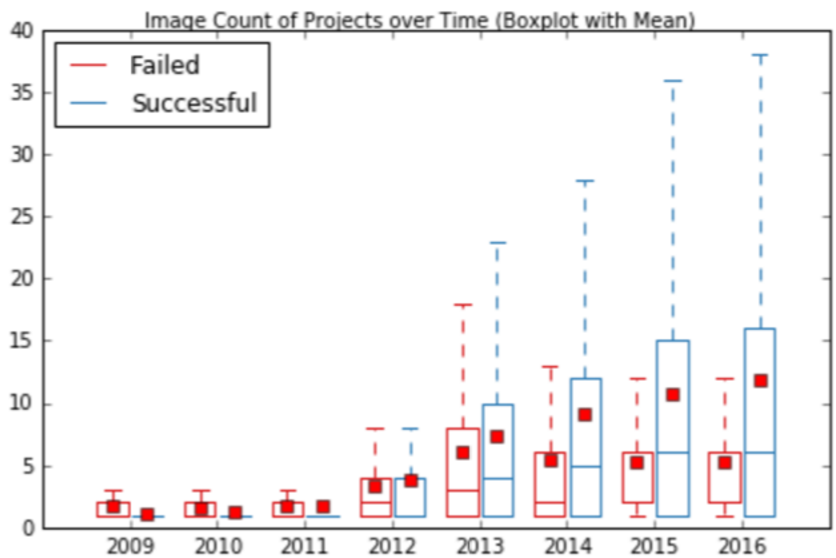

Figure 3: Boxplot for image counts, with mean indicated with red rectangle

This has indicated that the data from Kickstarter campaigns should be considered panel, rather than cross-sectional. The investigation of the impact of images should be limited to images from a short period of time, or panel regression should be used (Kuppuswamy & Bayus, 2015).

6. Timeline for Next Semester

Tasks	Time
Wrapping Naptha/adapting Naptha source code for local Terrasect OCR Engine	**Before W1**
Configuration of OCR on server	**W1 – W2**
Spell corrector – training image dataset	**W1 – W2**
Spell corrector - training	**W2 – W3**
Image classification	**W3 – W4**
Text extraction from images for projects in 2015 and 2016	**W4 – W6**
Text correction and running Lasso model Results interpretation	**W4 – W5**
Text extraction from images for projects from 2009 to 2014	**W5 – W6**
Text correction and running Lasso model; Results interpretation and comparison	**W5 – W7**

7. References

Abak, A. T., Baris, U., & Sankur, B. (1997). The performance evaluation of thresholding algorithms for optical character recognition. In*Document Analysis and Recognition, 1997., Proceedings of the Fourth International Conference on* (Vol. 2, pp. 697-700). IEEE.

Agrawal, A., Catalini, C., & Goldfarb, A. (2015). Crowdfunding: Geography, social networks, and the timing of investment decisions. *Journal of Economics & Management Strategy*, *24*(2), 253-274.

Beier, M., & Wagner, K. (2015). Crowdfunding Success: A Perspective from Social Media and E-Commerce.

Belleflamme, P., Lambert, T., & Schwienbacher, A. (2014). Crowdfunding: Tapping the right crowd. *Journal of Business Venturing*, *29*(5), 585-609.

Breuel, T. M. (2008). The OCRopus open source OCR system. In*Electronic Imaging 2008* (pp. 68150F-68150F). International Society for Optics and Photonics.

Cocciolo, A. (2015). The Rise and Fall of Text on the Web: A Quantitative Study of Web Archives. *Information Research: An International Electronic Journal*, *20*(3), n3.

Epshtein, B., Ofek, E., & Wexler, Y. (2010). Detecting text in natural scenes with stroke width transform. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on* (pp. 2963-2970). IEEE.

Lam, P. T., & Law, A. O. (2016). Crowdfunding for renewable and sustainable energy projects: An exploratory case study approach. *Renewable and Sustainable Energy Reviews, 60*, 11-20.

Li, Z., & Jarvenpaa, S. (2015). Motivating IT-Mediated Crowds: The Effect of Goal Setting on Project Performance in Online Crowdfunding. *Available at SSRN, 2672056.*

Marom, D., & Sade, O. (2013). Are the Life and Death of a Young Start-up Indeed in the Power of the Tongue? Lessons from Online Crowdfunding Pitches. *Available at SSRN, 2255707*.

Mitra, T., & Gilbert, E. (2014). The language that gets people to give: Phrases that predict success on kickstarter. Paper presented at the 49-61. doi:10.1145/2531602.2531656

Mori, S., Suen, C. Y., & Yamamoto, K. (1992). Historical review of OCR research and development. *Proceedings of the IEEE, 80*(7), 1029-1058.

Moritz, A., & Block, J. H. (2016). Crowdfunding: A literature review and research directions. In *Crowdfunding in europe: State of the art in theory and practice* (pp. 25-53). Cham: Springer International Publishing.

Kickstarter (2016). Kickstarter Stats. Retrieved from https://www.kickstarter.com/help/stats (accessed on October 24, 2016).

Kuppuswamy, V., & Bayus, B. L. (2015). Crowdfunding creative ideas: The dynamics of project backers in Kickstarter. *UNC Kenan-Flagler Research Paper*, (2013-15).

Kwok, K. (2014). Project Naptha. Retrieved from http://projectnaptha.com

Sas, J., & Zolnierek, A. (2013). Three-stage method of text region extraction from diagram raster images. In *Proceedings of the 8th International Conference on Computer Recognition Systems CORES 2013* (pp. 527-538). Springer International Publishing.

Schniter, E., Sheremeta, R. M., & Shields, T. W. (2015). The Problem with All-or-nothing Trust Games: What Others Choose Not to Do Matters In Trust-based Exchange. *Available at SSRN, 2708869*.

Smith, R. (2007). An overview of the Tesseract OCR engine.

Smith, S., Windmeijer, F., & Wright, E. (2015). Peer effects in charitable giving: Evidence from the (running) field. *The Economic Journal*, *125*(585), 1053-1071.

Sun, P. C., & Cheng, H. K. (2007). The design of instructional multimedia in e-Learning: A Media Richness Theory-based approach. *Computers & Education*, *49*(3), 662-676.

Thies, F., Wessel, M., Rudolph, J., & Benlian, A. (2016). Personality Matters: How Signaling Personality Traits Can Influence the Adoption and Diffusion of Crowdfunding Campaigns.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso.*Journal of the Royal Statistical Society. Series B (Methodological)*, 267-288

Wang, Z., Yang, J., Jin, H., Shechtman, E., Agarwala, A., Brandt, J., & Huang, T. S. (2015). DeepFont: Identify Your Font from An Image. In *Proceedings of the 23rd ACM international conference on Multimedia* (pp. 451-459). ACM.

Xiao, S., Tan, X., Dong, M., & Qi, J. (2014). How to Design Your Project in the Online Crowdfunding Market? Evidence from Kickstarter.

Zhang, J., & Liu, P. (2012). Rational herding in microloan markets. *Management science, 58*(5), 892-912.

Zhong, Y., Karu, K., & Jain, A. K. (1995). Locating text in complex color images. In *Document Analysis and Recognition, 1995., Proceedings of the Third International Conference on* (Vol. 1, pp. 146-149). IEEE.