Analyzing Persuasiveness through Multimodal Model

Ziming He
Georgia Institute of Technology
North Ave NW, Atlanta, GA 30332
zhe66@gatech.edu

Marcus Loo
Georgia Institute of Technology
North Ave NW, Atlanta, GA 30332
mloo3@gatech.edu

Abstract

Kickstarter is a platform that allows users to fund their projects through crowdfunding. Users decide what projects to crowdfund primarily by looking at the text and images the creator of the project put up. In this paper we analyze the efficacy of Kickstarter pages by using deep learning models to classify the likelihood of a project reaching its funding goals. We scraped 678 Kickstarter pages extracting their text and images. We employed multiple traditional machine learning models and pretrained deep learning models for image and text classification problems. The accuracy of the models seems to vary between models. We also provide possible causes for the accuracy and potential future work directions.

1. Introduction

Persuasion is a powerful tool that allows for change. This change can come in the many forms. It can be through something complicated like laws, or something as simple as funding for a personal project. Kickstarter is the largest crowdfunding platform online that can help with the latter. On Kickstarter, 17 million people pledged over $4.5 billion to over 450 thousand Kickstarter projects, but unfortunately, only 37.44% projects succeeded in getting the necessary funding [1]. We want to better understand how the roughly 150 thousand successfully funded projects were able to persuade backers to pledge to their projects, and why 290 thousand projects were unable to get the funding they needed.

A Kickstarter campaign is generally comprised of 3 parts: a video, text, and images. As a result of limited resources, we have decided to only focus on two of these aspects - text and images.

1.1. The Dataset

The dataset that we used consist of metadata such as project name, project category, main category, currency, deadline, goal, launched, pledged, backers, country, usd pledged. All the metadata are publicly available at Kaggle. In addition, we crawled additional information from the project webpages that includes at least one image and one line of description that we call, the blurb.

1.2. Dataset Category

Here are some descriptions for each category of data that we used.

- Project Name: The name of the project. We used this as an identifier in crawling and labeling the each sample
- Project Category: The subcategory that the project belongs to. This includes Poetry, Food, Documentary, Comic books, etc.
- Main Category: The main category that the project belongs to. This includes Comic, Film & Video, Music, Food, etc.
- Currency: The currency that the project pledged in. This includes USD, GBP, etc.
- Deadline: The deadline for the project to be successfully pledged. The date is in the format of YYYY/MM/DD HH:MM:SS.
- Goal: The amount of funding the project asked for. This amount is in the unit of currency mentioned above.
- Launched: The date that the project launched its campaign. The date is in the format of YYYY/MM/DD HH:MM:SS.
- Pledged: The amount that the crowd gave to the project. The unit is also in the currency mentioned above.
- Backers: The number of backers who supported the project by donating money to the project.

- Country: The origin country that the campaign launched. The two major categories include US (80%) and Great Britain (8%).
- USD Pledged: The total amount that the crowd donated translated in USD using the rate on the date of crawling.
- Image: An image that gets displayed when users browse through the Kickstarter website.
- Description (blurb): A short description that the users sees before they decide to go to the main webpage of the project. This description is usually one line or two.

1.3. Relevant Work

Analyzing persuasiveness has been an large focus of natural language processing. Throughout the years, researchers have been developing new models to analyze the techniques and signals that are present in the persuasive texts. Usually this type of analysis is done on websites similar to Kickstarter, such as Advertisement and Loaning.

Dey et al.[3] analyzed the persuasiveness of Kickstarter project by conducting human evaluation using Amazon MTurk on videos they crawled from the 210 Kickstarter projects. They discovered that different factors contribute differently across various categories. For example, in fashion and design category, more attention is paid towards the appearance of the presenter, whereas in technology category, more attention is paid towards the usefulness or utility of the project. Etter et al.[4] also discovered that project with a higher social media presence also tend to be more successful. Kickstarter users prefer to be constantly updated on the project they backed. And the users tend to fund projects which is launched by someone they previously backed with success. They also tend to back similar projects. This effect is known as peripheral persuasiveness when people are not persuaded by the argument or advertisement itself, but also by his or her own prior experience. Mitra et al.[9] analyzed keywords and sermantics in the Kickstarter projects. They discovered three major traits of a successful Kickstarter campaign: reciprocity, social proof, and expert opinions. Reciprocity means that people tend to get persuaded by an advertisement or argument if they are hinted or promised favors to be returned. They will more likely fund a project if they are promised a bigger return. Social proof means that users tend to agree more often with people who reveals their identity and they will more likely donate to a campaign which there were already a large group of backers. Expert opinions means that the user are often affected by the opinion of other backers who have been actively funding campaigns. They are also more likely to be persuaded by a project if the person behind the project displays certain level of expertise in fulfilling and delivering the project.

In similar approaches, Yang et al.[16] analyzed the effectiveness of persuasion on an online forum called Kiva, where users pledge other users to loan to their personal causes such as food needs and medical bills. They investigated the effect of interpersonal relations on persuasion. A higher level of activities between team members and the group leader often leads to stronger support from the group leader, thus making their arguments more personal and more effective to the audience. Frankz et al.[5] analyze the persuasiveness of political advertisements. They determined that with all else equal, more exposure to one candidate's advertisement will often result in a more favorable outcome for the candidate.

1.4. Motivation

Although there have been fair amounts of researches focusing on analyzing traits that contribute to the effectiveness of arguments on online funding forum like Kickstarter, Kiva, and GoFundMe. Majority of the researches only focus on one modality of the data, such as text, video, and images. Our work focuses on analyzing the effectiveness based on multiple signals, combining images with texts, which is the predominate version of advertisements we see online nowadays. We hope that our work can be the starting point of more analysis into applying multi-modal model to analysis of persuasion or other quality of multimedia contents that are present everywhere online and in real life.

2. Approach

In the simplest sense, the problem of persuasiveness of advertisement is a binary classification problem. The inputs include all the metadata mentioned in Section 1.2, the text description, and an image. The output is either success or failure. We also classify canceled projects as failed. Hence, the output will be binary instead of ternary. We employed a few different models to predict the success of the Kickstarter campaign. In addition, we wish that we can also compare the results of deep learning models vs traditional machine learning models. Therefore, for baseline models, we employed models such as Random Forest Classifier (RF) and Support Vector Machine (SVM). For deep learning models, we break it down into three individual sections, using Convolutional Neural Networks (CNN) for classifying the image inputs, Bidirectional Transformers for classifying the text inputs, and combining CNN and Bidirectional Transformers as a multimodal model to experiment on possible advantages of combining multiple inputs for classification. For all the models, we have a 80/20 split between train and test data. We used Pytorch, Keras and Sklearn to build the various models [10, 2, 11].

2.1. Baseline Models

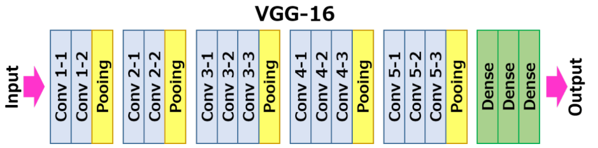
The baseline models that we employed are random forest classifier and support vector machine. For the inputs, we resize the image to be 28x28 and use bag-of-words to preprocess the text inputs. We then flatten the input image to be a size 784 input vector, concatenate that with the bag-of-words vector and the metadata input vector to be fed into the RF and SVM. Random Forest is an ensemble learner composed of various weak learners that generally works well in a group. However, Random Forest is highly uninterpretable. Support Vector Machine is vastly used in binary classification problem, it is robust to large number of variables and low amount of data samples. The base models all use cross-entropy loss and Adam optimizer [7].

2.2. VGG16 for Images

For analyzing the image inputs, we decided to use the classical VGG architecture which has proven to be a robust model that is often used in tasks such as classification, caption generation, semantic segmentation, and etc [14]. In particular, we use the VGG16 architecture which has 16 layers. In order to save time for training because by the end we will be combining the CNN models, we also imported pretrained weights of VGG16 architecture on the ImageNet dataset.

From the Kickstarter data that we crawled, there is at least one image that is present in each project. It is the thumbnail of the project that Kickstarter sees when they browse through different projects. All the images that we crawled have the same size, but we still resize all the images to be 224x224 so they fits the input parameter shape for the VGG architecture. We used binary cross entropy loss for the model, and we reshape the last layer so the output is a single binary output. The CNN model uses ADAM optimizer [7].

Figure 1. The VGG16 CNN Architecture. Source: Neurohive



2.3. Bidirectional Transformers for Texts

The binary classification model that we decided to use was a deep bi-directional transformer. Transformers felt like a fitting model to test with as it is more parallelizable than Recurrent Neural Networks (RNNs) and Long short-term memory (LSTMs), which would allow for training with larger data a lot quicker than if using RNNs or LSTMs [15]. We also wanted to see how state-of-the-art architectures can help predict and classify persuasiveness. In order to speedup time, we decided to use a pretrained language model developed by Facebook called RoBERTa. This model is similar to Google's BERT, but has an increase to training size and hyperparameter tuning, which has generally improved accuracy [8].

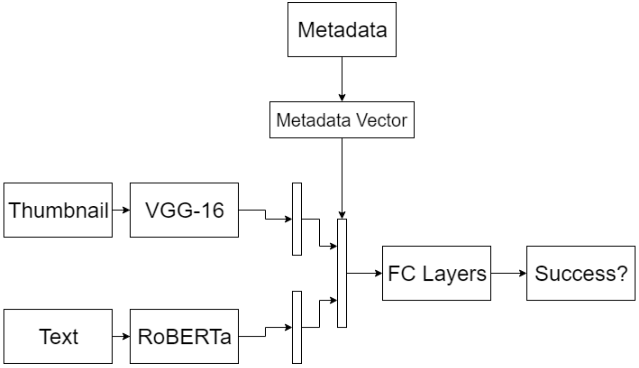
In the Kickstarter data that we scraped, every project had a blurb. There were also projects that had other texts including text pertaining to rewards and a more in-depth description. Initially we were going to employ a binary classification of using just the blurb; however, with initial testing, this proved to not result in any useful classification accuracy. We decided to use the other text parts as well in order to obtain more text information for better classification. The model that we are using cannot read text directly. BERT takes in a sequence of tokens as input. Therefore we tokenize the text, and use special tokens to represent the ends of sentences inside each wall of text. There is also a hyperparameter that determines the size of this sequence, which means that if the text is shorter than the determined size, it will be padded, and if the text is longer than this, text will be truncated.

This model uses ADAM optimizer [6]. As this is a classification problem, we decided to use cross-entropy loss as our loss function, and our output from our model will have a shape (number of samples, number of classes). The output can be interpreted by taking the softmax of the output, which would give the probabilities of a particular class. The class with the largest probability is then the models prediction for whether or not the text helped successfully a Kickstarter project.

2.4. Combining Images and Texts

In the very end, we decide to combine the image classification model and the text classification model by concatentating the outputs of the two models before the softmax layer to be a long vector and pass that long vector as an "input" to three fully connected layers as a classifer. The output of the last three fully connected layer is then passed into a softmax layer which outputs either 0 or 1 indicating the prediction of the campaign being successful or not.

Figure 2. The Multimodal Architecture



3. Experiments and Results

Here are the results of various models that we have tried. We did not have access to a GPU, therefore we were limited on the amount of training that we could do for our models. As a result, for VGG16, RoBERTa, and the Multimodal models, we used random search to try to tune parameters as well as use recommendations from other papers.

	Training Accuracy	Test Accuracy
RF	69.34%	62.36%
SVM	70.24%	67.34%
VGG16	53.14%	51.93%
RoBERTa	99.46%	79.46%
Multimodal	63.74%	59.18%

Most of the models do not seem to be overfitting. However, RoBERTa is overfitting quite a bit. This can be seen through its training accuracy performing a lot better than its testing accuracy. As seen above, traditional machine learning models actually outperforms deep learning architecture trained on only the thumbnails. One possible reason for this might be that the thumbnails in the some Kickstarter campaigns are stock images that are not visually strongly correlated with the projects themselves. Therefore, by using only the images, there is high possibility that the users do not know what kind of projects the campaign is funding. Out of all the models that we have tested, RoBERTa seems to have performed the best. Analyzing the results of the model, it surprisingly favored short and succinct descriptions. Looking at the texts it also seemed like a lot of the correctly guessed advertisements related to music or film. This seems to indicate that this model found correlations between certain genres and its likelihood of getting its necessary funding. Surprisingly, the multimodal deep learning model did not perform as well as expected, falling short of the baseline models. One possible reason might be that we froze all too many layers in VGG16 and RoBERTa, making it hard to learn features needed for predicting the success of the Kickstarter campaigns. Another possibility is that the vision pipeline does not work as well as the text pipeline so it brought down the accuracy when combining both models into a single multimodal model.

4. Applications and Future Work

Although the models that we tried, including the traditional machine learning models, achieves fairly reasonable results, there is still large room of improvements. For future work, we are looking into the possibility of improving the accuracy as well as making the results of the models more interpretable by humans. The dataset that we used to train and evaluate our models only had 678 samples. Increasing the amount of samples could allow our models to increase in accuracy as well as better reflect persuasiveness on Kickstarter. We are also looking into using attention mechanism on the vision part of the pipeline that process the image inputs so the model can be used to highlight both parts of images and texts that the model used to judge whether the campaign will be successful or not. Future works should also look into the possibility of using adversarial feature mining and the potential of controlling cofounders [12, 13]. We hope our work can help advertisement maker and crowdfunding individuals to make their arguments or advertisement more effective so that they can achieve their goal more easily.

References

[1] Kickstarter stats. 2019. [Online; accessed 3-Dec-2019].

[2] François Chollet et al. Keras. `https://keras.io`, 2015.

[3] Sanorita Dey, Brittany Duff, Karrie Karahalios, and Wai-Tat Fu. The art and science of persuasion: not all crowdfunding campaign videos are the same. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, pages 755–769. ACM, 2017.

[4] Vincent Etter, Matthias Grossglauser, and Patrick Thiran. Launch hard or go home! predicting the success of kickstarter campaigns. In *Proceedings of the first ACM conference on Online Social Networks (COSN'13)*, number CONF, pages 177–182. ACM, 2013.

[5] Michael M Franz and Travis N Ridout. Does political advertising persuade? *Political Behavior*, 29(4):465–491, 2007.

[6] Kenton Lee Kristina Toutanova Jacob Devlin, Ming-Wei Chang. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

[7] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[8] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*, 2019.

[9] Tanushree Mitra and Eric Gilbert. 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*, pages 49–61. ACM, 2014.

[10] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. Automatic differentiation in pytorch. 2017.

[11] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[12] Reid Pryzant, Sugato Basu, and Kazoo Sone. Interpretable neural architectures for attributing an ad's performance to its writing style. In *Proceedings of the 2018 EMNLP Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP*, pages 125–135, 2018.

[13] Reid Pryzant, Youngjoo Chung, and Dan Jurafsky. Predicting sales from the language of product descriptions. 2017.

[14] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[15] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in neural information processing systems*, pages 5998–6008, 2017.

[16] Diyi Yang and Robert E Kraut. Persuading teammates to give: Systematic versus heuristic cues for soliciting loans. *Proceedings of the ACM on Human-Computer Interaction*, 1(CSCW):114, 2017.