CROWDFUNDING SUCCESS PREDICTION USING PROJECT TITLE IMAGE AND CONVOLUTIONAL NEURAL NETWORK

Matko Šarić[1], * and Marija Šimić Šarić[2]

[1]University of Split – Faculty of Electrical Engineering, Mechanical Engineering and Naval Architecture
Split, Croatia
[2]University of Split – Faculty of Economics, Business and Tourism
Split, Croatia

DOI: 10.7906/indecs.21.6.8
Regular article

Received: 23 August 2022.
Accepted: 12 October 2023.

ABSTRACT

Prediction of crowdfunding success is a challenging problem that has great importance for project creators and platforms. Although meta features, e.g., number of updates or backers, are widely used for success prediction, they are limited to time period after project posting where project creators cannot adapt their profiles. Because of that, ability to predict campaign success in pre-posting phase would significantly improve chance for project success. According to the theory, mostly used methods in this situation are those based on text features, while methods based on the influence of image modality on project success are rare. Due to this, in this article we propose deep learning-based method for crowdfunding success prediction in pre-posting phase using project title image. Experimental results show that image modality could be used for campaign success prediction. Proposed method obtains results comparable to competing methods from literature, but using only one image per campaign and no derived features. It is also shown that deeper convolutional neural network achieves better prediction performance.

KEY WORDS

crowdfunding, success prediction, project title image, deep learning

CLASSIFICATION

JEL: O31

*Corresponding author, η: msaric@fesb.hr; +385 (21) 305 633;
FESB, Ul. Ruđera Boškovića 32, HR – 21 000 Split, Croatia

INTRODUCTION

The global crowdfunding market in 2020 was valued at 12,27 billion U.S. dollars [1] and consist of reward-based crowdfunding, equity – based crowdfunding, donation – based crowdfunding and real estate crowdfunding [2]. Crowdfunding success prediction, especially in phase before project posting, is a challenging task that has great importance for project creators as well as for crowdfunding platforms. Reliable prediction would allow project creators to revise campaign profile in timely manner and maximize chance for success, while crowdfunding platforms could emphasize projects having higher probability to reach their goal. Existing approaches for success prediction mainly exploit dynamic meta-data after the project is posted. In pre-posting phase focus is put on profile text, that is mainly project description. Although visual content carries more information content than text, influence of campaign images on final outcome has not been extensively studied.

Deep learning, as subset of machine learning, has been successfully applied in different areas like object detection in images [3], medical imaging [4], natural language processing [5], finance and banking [6] etc. Deep learning models consist of multiple processing layers that reveals hidden structures in high-dimensional data. Each added layer represents input data on more abstract level that is suitable for detection or classification tasks. Convolutional neural network (CNN) is deep learning architecture that has made breakthrough in image and video processing showing strong performance in image classification task.

In this article we propose method for crowdfunding success prediction based on project title image and convolutional neural network. Image dataset is collected by scraping project title images from Kickstarter platform. Using this dataset, we have trained CNN-based classifier that predicts whether project is successful or not using only project title image as input. Experimental results show that proposed method achieves state of the art results meaning that even single project title image has predictive potential that could be considered for success prediction, especially in synergy with textual and meta features. Different from other approaches in literature, we predict project success using automatically extracted CNN features of title image only. In this way we avoid manual selection of image features used for classification. We performed comparison of 3 widely used CNN architectures showing that deeper network achieves better results.

LITERATURE REVIEW

Methods for crowdfunding success prediction mostly utilizes machine learning techniques. Greenberg et al. [7] have trained decision tree classifier for project success prediction using meta features (project goals, sentence count, project duration etc.) where they achieved 68% accuracy. In [8] novel text analytics framework for crowdfunding success prediction is introduced. Authors developed model for extraction of topical features from project descriptions which are then combined with numerical features and used as an input of different classifiers. Evaluation showed that in prediction performance decision tree classifier outperforms SVM, backpropagation neural network and extreme learning machine. It is important to note that research was conducted on two popular crowdfunding websites (Dreamore and Zhongchou). Li et al. [9] applied censored regression for crowdfunding success prediction and showed that addition of temporal features obtained after project launch significantly improves prediction performance.

Little work has been done regarding influence of visual content on crowdfunding success. In [10] authors proposed multimodal representation of campaign including text, images and metadata. Textual features include: title, summary, project description and risks/challenges. Visual representation of single campaign consists of all jpg images from campaign website, while metadata include campaign category and project goal. These features are combined in 3 branch CNN architecture to predict project success. Bottom branch encodes text using Bag of Words (BoW) with Term Frequency-Inverse Document Frequency method (TF-IDF). Middle branch encodes campaign visual content utililizing ImageNet pre-trained 16-layer VGG model as feature extractor for each image in single profile. Final feature map for profile is created by stacking VGG features of single images. Top branch encodes metafeatures with fully connected neural network. Performance of success prediction with project images only is also investigated and it is shown that visual modality has significant contribution to campaign success, but to lesser extent than textual information.

Zhang et al. [11] combined textual and image modalities to predict crowdfunding campaign outcome for GoFundMe platform. Part of project features are crawled directly from websites (launch date, description, location, title cover image, category, current amount, goal amount etc.), while rest of features are derived (fundraisers location population, image quality, number of faces in images etc.). Different from [10], image content is not represented directly with CNN features, but with aesthetic and technical quality scores for title image as well as features derived from face recognition (number of faces, gender, beauty smile level, emotion, age). Image quality features are obtained with CNN model pretrained with ImageNet dataset and fine-tuned for classification of visual quality. Deep learning based face recognition platform Face++ is used to get facial features. Textual information is encoded using Linguistic Inquiry and Word Count (LIWC) features which are then fused with image features and classified using random forest classifier. Results showed that for projects belonging to 3 categories (Competitions & Pageants, Community & Neighbors, and Weddings & Honeymoons) and having goal between $ 8 000 and $ 40 000 image quality score, when fused with basic and text LIWC features, significantly improves classification performance. Interestingly, using image quality as single feature gives best classification result for this project group. It is also shown that for projects in other categories image quality does not influence on crowdfunding success. In [12] a machine learning approach is employed to recognize faces and facial expressions in profile images. It is found that appearance of smiling faces gives 5 % increase of funding amount, while presence of creator’s face negatively influences on it.

DATA AND METODOLOGY

Training and testing of proposed method have been performed on image dataset scraped from Kickstarter. More precisely, we used dataset in csv format available at Webrobots [13] containing campaigns in the period from June 2010 to February 2021. Links to campaign cover images are extracted from csv files and download was performed with script written in Python. Cover images for projects with states canceled, suspended and live were filtered out what finally gives 54 563 images for successful campaigns and 33 159 images for failed campaigns. Dataset is split in 3 parts: training, validation and testing dataset.

Overview of proposed method for crowdfunding success prediction is shown in Figure 1.

Campaign profile image is used as input of CNN that was previously trained on ImageNet [14] dataset that is widely used in visual object recognition research. Images are annotated as one of 1000 categories from ImageNet Large Scale Recognition Challenge (ILSVRC). Here we

Table 1. Sizes of datasets used for training, validation and testing.

Dataset	Number of images		
	Success	Fail	Total
Training	31 500	17 161	48 661
Validation	3 214	1786	5 000
Testing	19 849	14 212	34 061

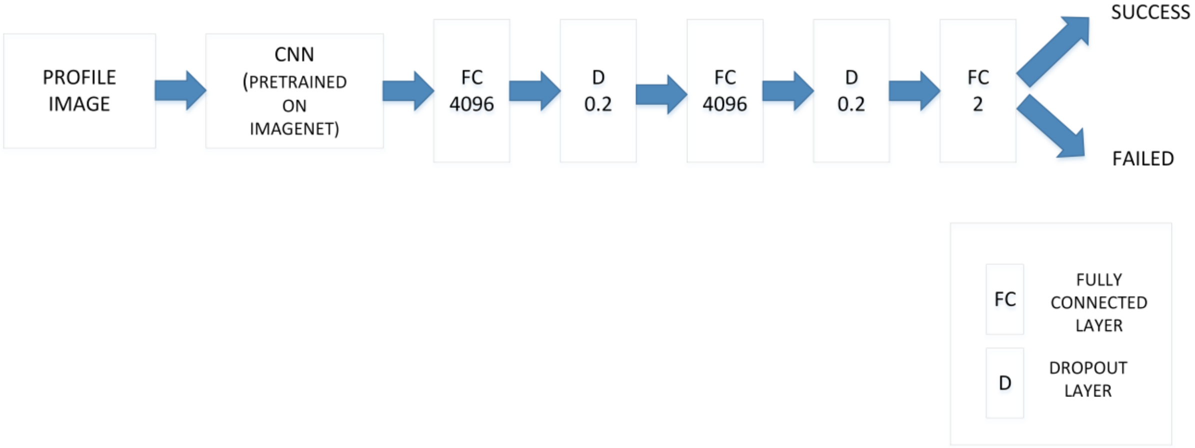

Figure 1. Overview of proposed method for crowdfunding success prediction.

tested 3 CNN architectures widely used in computer vison tasks: 16-layer VGG model [15], ResNet 50 [16] and DenseNet121 [17]. Top classification layer is removed from CNN and 2 fully connected layers with 4 096 channels are added followed by 2 sigmoid activated outputs representing campaign success or failure In this way transfer learning approach is realized where pretrained CNN acts as feature extractor followed by fully connected classifier. During training CNN weights are frozen and only weights from fully connected classifier are updated. In this way visual features learned from ImageNet dataset are repurposed for task of classification of campaign title image on success or failure classes. Since training the CNN from the scratch requires significant hardware resources and large training dataset, transfer learning allows us to perform training faster and with limited number of training images. We hypothesize that CNN features learnt on ImageNet dataset can be exploited for crowdfunding success prediction, although connection between project image content and project success/failure is far from straightforward and it is more complex than classification of image into classes like “horse”, “car”, “tree” etc.

VGG architecture consists of convolutional layers that use 3×3 kernels giving relatively small receptive field. Max pooling layers downsamples feature maps by factor 2. VGG network is shown in Figure 2.

Increasing the CNN depth by adding more convolutional layers enables extraction of high level features that help network to learn complex mapping between input (profile image) and output (success or failure). Problem here is that simple stacking more layers leads to accuracy degradation. This can be handled by ResNet architecture [16] where residual mapping is fitted instead of direct mapping between input and output. Residual block shown in Figure 3 learns residual mapping $F(x) = H(x) - x$, where x is input and $H(x)$ is original mapping. Huang et al. [17] proposed DenseNet architecture where all layers with same feature size are connected with each other. Each layer receives inputs from all preceding layers and passes its feature maps to all subsequent layers. If x_l is feature map of l^{th} layer and $x_0, x_1, \dots, x_{l-1}$ are feature maps of all preceeding layers with same size, then

$$x_l = H_l([x_0, x_1, \dots, x_{l-1}]), \tag{1}$$

where $[x_0, x_1, \dots, x_{l-1}]$ represents concatenation of feature maps and H_l is non-linear transformation of layer l composed of and batch normalization (BN), rectified linear units (ReLU) and convolutional operations. Since concatenation is viable only for feature maps of the same size, there is no pooling operation in function H_l which is important part that enables features downsampling in CNN. Therefore, network is divided in dense blocks (H_l) and transition layers between them that performs convolution and pooling, Figure 4.

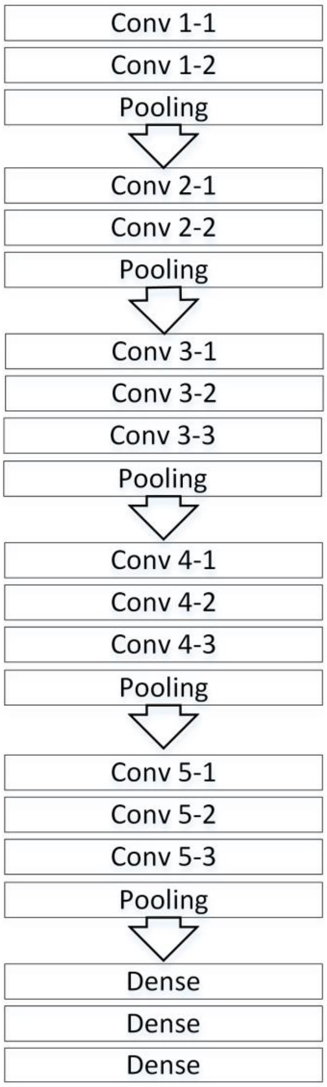

Figure 2. VGG architecture.

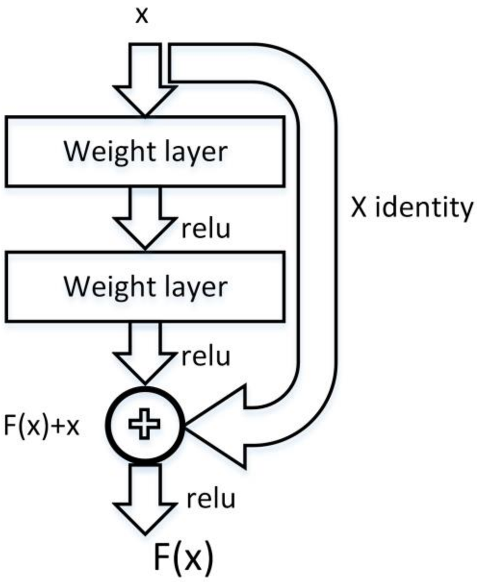

Figure 3. ResNet architecture.

Since we deal with binary classification problem, binary cross-entropy is chosen as loss function:

$$L = -\sum_{k=1}^{N}[\delta(y_k = 1)\log p_k + \delta(y_k = 0)\log(1 - p_k)], \tag{2}$$

where δ is indicator function having value 1 when prediction corresponds to ground truth (otherwise it has value 0), $p_k \epsilon [0, 1]$ is estimated probability for class with label 1 that is obtained as output of last dense layer with sigmoid activation.

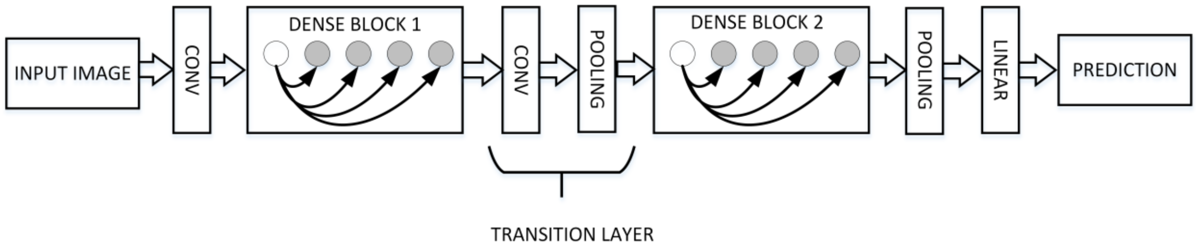

Figure 4. DenseNet architecture with 2 dense blocks.

RESULTS AND DISCUSSION

For performance evaluation and comparison with competing methods we have used following metrics: Accuracy, Recall, Precision, F-score and AUC@ROC. Accuracy is defined as:

$$Accuracy = \frac{Number\ of\ correct\ predictions}{Total\ number\ of\ predictions}. \tag{1}$$

If binary classification is considered, accuracy can be calculated with

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}, \tag{2}$$

where TP refers to the number of true positive cases, TN is number of true negative, FP is number of false positive and FN is number of false negative outputs. Recall is the measure that reveals proportion of true positives that are detected correctly:

$$Recall = \frac{TP}{TP+FN}. \tag{3}$$

Precision shows what proportion of positive identification is actually correct, that is proportion of false positive outputs:

$$Precision = \frac{TN}{TN+FP}. \tag{4}$$

F-score is the measure defined as

$$F - score = 2\frac{Precision * Recall}{Preision + Recall}. \tag{5}$$

This can be interpreted as weighted average of precision and recall.

Receiver operating characteristics (ROC) curve plots relation of true positive rate and false positive rate. Area under curve (AUC) is defined as area under ROC curve describing how classification performs over range of all classification thresholds. Training is performed using Python and Keras library with TensorFlow backend. We used SGD optimizer and learning rate set to 1e-4 for 20 epochs. Batch size was 8 and early stopping is employed if validation loss did not decrease for 6 epochs.

Table 2. Results of success prediction for different CNN architectures.

Classifier		Accuracy	Precision	Recall	F-score	AUC
VGGNet16		0,6139	0,6141	0,9083	0,7328	0,6452
ResNet50		0,6431	0,6545	0,8212	0,7284	0,6803
DenseNet121		0,6502	0,6568	0,8373	0,7361	0,6890

Table 2 shows success prediction performance for VGGNet, ResNet and DenseNet classifier. It can be seen that DenseNet arhitecture achieves highest accuracy, precision F-score and AUC. ResNet predictor performs slightly worse for all measures. VGGNet architecture gives lowest accuracy, precision and AUC, but it has highest recall among proposed methods. This indicates that it can predict success of campaign (true positive case) with high reliability, while low precision means that classifier gives less reliable prediction of campaign failure. It can be seen that architectures with more layers performs better. This could be explained by influence of CNN depth, that is number of layers, where deeper architectures represent input image on higher abstraction levels needed to model non-trivial relation between profile image and campaign success or failure. Drawback of deep architectures is higher number of network parameters what implies larger training set to avoid overfitting. This problem is overcome by using pretrained weights and fine-tuning only the last two fully connected layers.

Regarding competing methods from literature, proposed approach could be compared with method presented in [10] where visual features are also used for campaign success prediction as single modality and in combination with textual and meta features. Key difference in comparison with proposed approach is that in [10] all campaign profile images (with size greater than 200 pixels) were used as input of CNN. For each image, its feature map was extracted using pretrained VGGnet16 model. Aggregated feature maps of all images are used as CNN input. In contrast, we used only one (title) image as input to CNN classifier.

Performance comparison is given in Table 3. Our method has significantly higher recall and better F-score, while precision and AUC values are lower compared to results from [10]. Overall, we obtain comparable performance with higher recall, but representing campaign with only one profile image.

Table 3. Results comparison with competing method [10].

Method	Precision	Recall	F-score	AUC
Proposed method (DenseNet121)	0,6568	0,8373	0,7361	0,689
Cheng [10] (visual modality only)	0,6809	0,6738	0,6768	0,7340
Zhang [11] (derived image features, projects with goal $ 8 000-$ 40 000)	0,88	0,83	0,81	NA

Regarding other methods from literature, it is hard to make direct comparison because there has not been done much work with success prediction using image content. In [11] authors investigated influence of image modality on campaign success for GoFundMe platform. Project title image is represented with aesthetical and technical scores obtained with pretrained deep learning model. Also, facial features extracted with Face++ recognition platform (number of faces, gender, beauty etc.) are added to investigate the influence of facial attributes. Random Forest is chosen as classifier and evaluation is performed for case when image quality features are used as input to classifier. For each project category authors used different image quality features. Classification with image quality gives higher precision and recall comparable to our method, but only for campaigns with goal between $ 8 000 and $ 40 000.

It should be noted that our analysis is conducted for all campaigns regardless target amount. Also, important difference is that we represent campaign with CNN features of title image without derived visual features. In this way we avoid manual selection of image features for each category moving this task to CNN.

CONCLUSION

In this article we deal with problem of crowdfunding campaign success prediction in pre-posting phase exploring the influence of visual modality on final outcome. Experiments are performed with image dataset scraped from Kickstarter. Three widely used CNN architectures (VGG16, ResNet50, DenseNet121) are used as feature extractors followed by fully connected neural network as binary classifier. Evaluation shows that DenseNet121 CNN architecture has performance comparable to state-of-the-art methods. Different to other methods, our approach uses only one image per campaign and no derived fetaures leaving the task of feature selection to CNN. Future work would include addition of text and meta features to build stronger multimodal campaign success predictor.

REFERENCES

[1] –: *Statista - global crowdfunding market size*.
https://www.statista.com/statistics/1078273/global-crowdfunding-market-size, accessed March 2022,

[2] –: *Statista - Europe alternative finance transactions crowdfunding*.
https://www.statista.com/statistics/412487/europe-alternative-finance-transactions-crowdfunding, accessed March 2022,

[3] Liu, L., et al.: *Deep Learning for Generic Object Detection: A Survey*.
International Journal of Computer Vision **128**(2), 261-318, 2020,
http://dx.doi.org/10.1007/s11263-019-01247-4,

[4] Esteva, A., et al.: *Dermatologist-level classification of skin cancer with deep neural networks*.
Nature **542**(7639), 115-118, 2017,
http://dx.doi.org/10.1038/nature21056,

[5] Otter, D.W.; Medina, J.R. and Kalita, J.K.: *A Survey of the Usages of Deep Learning for Natural Language Processing*.
IEEE Transactions on Neural Networks and Learning Systems **32**(2), 604-624, 2020,
http://dx.doi.org/10.1109/TNNLS.2020.2979670,

[6] Huang, J.; Chai, J. and Cho, S.: *Deep learning in finance and banking: A literature review and classification*.
Frontiers of Business Research in China **14**, No. 13, 2020,
http://dx.doi.org/10.1186/s11782-020-00082-6,

[7] Greenberg, M.D.; Pardo, B.; Hariharan, K. and Gerber, E.: *Crowdfunding support tools: predicting success & failure*.
In: Mackay, W.E.; Brewster, S. and Bødker, S., eds.: *CHI'13 Extended Abstracts on Human Factors in Computing Systems*. IEEE, pp.1815-1820, 2013,
http://dx.doi.org/10.1145/2468356.2468682,

[8] Yuan, H.; Lau, R.Y.K. and Xu, W.: *The determinants of crowdfunding success: A semantic text analytics approach*.
Decision Support Systems **91**, 66-76, 2016,
http://dx.doi.org/10.1016/j.dss.2016.08.001,

[9] Li, Y.; Rakesh, V. and Reddy, C.K.: *Project success prediction in crowdfunding environments*.
In: Bennet, P.N. and Josifovski, V., eds.: *Proceedings of the Ninth ACM International Conference on Web Search and Data Mining*. ACM, New York, pp.247-256, 2016,
http://dx.doi.org/10.1145/2835776.2835791,

[10] Cheng, C.; Tan, F.; Hou, X. and Wei, Z.: *Success Prediction on Crowdfunding with Multimodal Deep Learning*.
In: Kraus, E., ed.: *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence*. IJCAI, pp.2158-2164, 2019,
http://dx.doi.org/10.24963/ijcai.2019/299,

[11] Zhang, X.; Lyu, H. and Luo, J.: *What Contributes to a Crowdfunding Campaign's Success? Evidence and Analyses from GoFundMe Data*.
arXiv preprint arXiv:2001.05446v3[cs.SI],
http://dx.doi.org/10.48550/arXiv.2001.05446,

[12] Kim, J. and Park, J.: *Does facial expression matter even online? An empirical analysis of facial expression of emotion and crowdfunding success*.
In ICIS 2017 Proceedings, 2017,

[13] –: *Web Robots*.
https://webrobots.io/kickstarter-datasets, accessed February 2021,

[14] –: *ImageNet*.
http://www.image-net.org, accessed 29 March 2021,

[15] Simonyan, K. and Zisserman, A.: *Very deep convolutional networks for large-scale image recognition*.
Preprint. arXiv:1409.1556v6[cs.CV],
http://dx.doi.org/10.48550/arXiv.1409.1556,

[16] He, K.; Zhang, X.; Ren, S. and Sun, J.: *Deep residual learning for image recognition*.
In: Agapito, L., et al., eds.: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2016,
http://dx.doi.org/10.1109/CVPR.2016.90,

[17] Huang, G.; Liu, Z.; Van Der Maaten, L. and Weinberger, K.Q.: *Densely Connected Convolutional Networks*.
In: Liu, Y., et al., eds.: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2017,
http://dx.doi.org/10.1109/CVPR.2017.243.