Aug 10th, 12:00 AM

Reward-based Crowdfunding Success Prediction with Multimodal Data

Liqian Bao
University of Wisconsin Milwaukee, lbao@uwm.edu

Zongxi Liu
University of Wisconsin Milwaukee, zongxi@uwm.edu

Huimin Zhao
University of Wisconsin Milwaukee, hzhao@uwm.edu

Reward-based Crowdfunding Success Prediction with Multimodal Data

Completed Research

Liqian Bao
University of Wisconsin-Milwaukee
lbao@uwm.edu

Zongxi Liu
University of Wisconsin-Milwaukee
zongxi@uwm.edu

Huimin Zhao
University of Wisconsin-Milwaukee
hzhao@uwm.edu

Abstract

As an increasing number of crowdfunding platforms recommend that entrepreneurs post multimodal data to improve data diversity and attract investors' attention, it becomes necessary to study how functions of multimodal data take effect to predict fundraising outcomes (i.e., success or failure). There is a lack of research providing a comprehensive investigation of multimodal data in crowdfunding. Rooted in language and visual image metafunctional theories, we propose a framework to explore ideational, interpersonal, and textual metafunctions of multimodal data. We empirically examine the effectiveness of each metafunction, each modality, and their combination in predicting fundraising outcomes. The empirical evaluation shows the predictive utility of any metafunctions and metafunction combinations. The results also demonstrate that adding data modalities can help to improve the prediction performance.

Keywords

Fundraising success prediction; multimodal data; metafunction; visual features.

Introduction

Multimodality integrates multiple data modalities, such as linguistic, visual, gesture, color, design, and sound signals, to express the idea (Norris and Maier, 2014). With the development of artificial intelligence and big data, multimodality becomes one of the popular research areas of Information Systems since multiple modalities can provide complementary information and improve the performance of the overall decision-making process.

Crowdfunding platforms, as a form of online microfinance, allow entrepreneurs to display multimodal data to improve data diversity and attract investors' attention. However, past research has mainly studied the effects of description texts in crowdfunding (Zhou et al., 2018; Mollick, 2014), largely overlooking how multimodal data (e.g., texts and images) interact and influence the fundraising outcome. In an online environment, visual features can increase the credibility of texts and induce emotional appeal for readers (Garrett, 2002), possibly leading to favorable behaviors. We posit that multimodal representations, including textual and visual features, may improve the effectiveness of fundraising success prediction.

To gain a comprehensive view of linguistic and visual features of texts and images, we propose a framework for analyzing multimodality of crowdfunding by adopting the elements of three earlier frameworks specifically designed for analyzing languages and images: Halliday's metafunctions framework of languages (1985), Kress and Van Leeuwen's functional visual design (1996), and Royce's intersemiotic complementarity of languages and visual images (1998). Building on these theories, we formulate the metafunctions of multimodal data in a reward-based crowdfunding platform, to investigate the following research questions: (1) whether each metafunction of each data modality is valuable for predicting fundraising success; (2) whether data multimodality improves the prediction performance over a single

modality in terms of each metafunction; (3) whether the combination of metafunctions improves the prediction performance over a single metafunction.

Literature review

Systemic Functional Linguistic and Metafunctions

Systemic Functional Linguistic (SFL) was first developed by linguist Michael Halliday for teaching Mandarin in the early 1960s (Halliday, 1985) and then extended to the English language (Halliday, 1994). It provides the central theoretical framework for systemic functional approaches to multimodality to analyze the function and meaning of semiotic resources (O'Halloran, 2008a). Semiotic resources are theorized as realizing three different meaning functions (known as metafunctions). This meta-functional system is used to interpret how semiotic resources simultaneously construct experiences and logic (ideational meaning), enact social relations (interpersonal meaning), and organize a structured text (textual meaning). Drawing on insights of Halliday, researchers gradually extended the systemic function theory to non-verbal semiotic resources and media. Kress & van Leeuwen (1996) argued that the same metafunctions can be identified in visual resources. Ideational meaning, interpersonal meaning, and compositional (textual) meaning are applied in visual imagery.

Due to its comprehensive functional representation of language meaning, SFL becomes the fundamental theory to capture valuable features in information systems areas, including computer-mediated communication (Abbasi and Chen, 2008), social media (Dong et al., 2018), and tacit knowledge elicitation (Zappavigna and Patrick, 2010). However, these prior studies refer to SFL to extract text features only, overlooking its power in dealing with multimodal data. Moreover, according to our review of related studies, no work has been done to analyze crowdfunding outcomes rooted in metafunctions theory. Since crowdfunding websites are providing multiple semiotic modes, we propose to manipulate metafunction representations of multimodal data in crowdfunding to enhance fundraising success prediction.

Multimodality in Crowdfunding

Crowdfunding campaign entrepreneurs rely on multimodal data, including texts and images, to communicate the novelty and value of their ideas to backers (Yang et al., 2020). However, only a few studies have analyzed the multimodality of crowdfunding projects. Hou et al (2019) deployed LIWC (Linguistic Inquiry and Word Count) to extract emotions from description texts and applied a deep learning method to extract the emotion features from title images. Cheng et al (2019) applied Bag of Words (BoW) and word embedding (GloVe) to represent textual features and used a pre-trained VGG-16 model to extract visual features from crowdfunding images. Perez et al (2020) extracted sentiment, word importance, and named entity from description texts and reproposed a pre-trained ResNet-152 model to extract emotion (eight types, e.g., sadness, fear, amusement), appearance (a color or an object), and semantic (the logit presence of predetermined objects in each image) features from images to identify fraud in crowdfunding projects. Kaminski and Hopp (2020) used Google API to extract spoken words and appeared objects from speech and video to predict fundraising outcome.

To our knowledge, there is a lack of research providing a comprehensive investigation of multimodal data in crowdfunding. Prior studies have only considered one or two metafunctions of texts or images and ignored the interactions among the multiple modalities and the interactions among the three metafunctions. To fill the gaps, we propose a framework and explore features representing the ideational, interpersonal, and textual metafunctions of multimodal data in crowdfunding. We conduct several experiments to study the effectiveness of each metafunction, each modality, and their interactions in predicting fundraising outcomes.

Framework

The Ideational Metafunction Representation

The ideational metafunction of language shows how we represent experience in the language (Halliday, 1985), including the experiential meaning and logical meaning between clauses. It is concerned with the analysis of the sequence of parts (i.e., words, word groups, clauses, clause complexes, and paragraphs),

which develops the texts (O'Halloran, 2008b). Royce (1998) adopted Halliday's ideational meaning in representing the visual structures. He derived the represented participants, which correlate to the ideational metafunction. They are all the elements or entities that are actually present in the visual, whether animate or inanimate (Royce, 2013).

Deep neural networks are well known for their effectiveness in extracting information from a large and unstructured dataset. They have been used for implementing a universal learning approach in different application domains (speech, language, and vision understanding) (Alom et al., 2019). We applied transfer learning using a popular language representation model BERT (Bidirectional Encoder Representation from Transformers) (Devlin et al., 2018) to present text ideational metafunction. Specifically, we processed texts into input embeddings (token embeddings, segment embeddings, and position embeddings) required by BERT and then initialized the BERT model with the pre-trained parameters (https://github.com/google-research/bert/blob/master/multilingual.md). We used the textual embedding outputs as the input features of prediction models.

Krizhevsky et al. (2012) made a breakthrough for image classification using convolutional neural networks (CNNs), which can significantly improve the description capability of image representation (Liu et al. 2018). We applied the VGG-16 architecture (Simonyan and Zisserman, 2014), which has been shown to achieve better recognition or classification accuracy in CNN (Alom et al., 2019), to map images into deep representations (https://github.com/minar09/VGG16-PyTorch).

Therefore, we obtained text embeddings and image embeddings from deep neural network models to represent the ideational meanings of crowdfunding projects.

The Interpersonal Metafunction Representation

The interpersonal metafunction is realized by the clause as an exchange of information or exchange of goods and services and is basically concerned with enacting social relationships between the speaker or writer and the audience or viewer in a specific context of communication (Halliday, 2004). Interpersonal meaning includes the forms of interaction and social interplay with others, polarity (positive and negative), and modality1 (degree of certainty and probability) (Halliday, 2004; Guijarro, 2010). Therefore, we deployed LIWC (Tausczik and Pennebaker, 2010) to obtain social relation (family, friends, females, males), polarity (positive and negative emotions), and modality (certainty, tentativeness) of each project description text as interpersonal metafunction representation of the text.

The interpersonal metafunction of visual images involves features of contact, social distance, and modality between viewers and visual participants. Contact is constructed by any gaze or facial expression of the visual participants to viewers, and it represents offering information in the form of a portrayal; social distance is determined by how close the visual participant appears to the viewer in an image, with close or long shots related to the degree of intimacy between visual participants and viewers (Kress and van Leeuwen, 2006). Modality is interpreted as the truth, credibility, and probability of what visual participants represent to viewers, and whether the information they offer is real or unreal (Royce, 2013). Pages that are relatively static, ordered, and less varied in color tend to have a higher modality and are more likely to be factual (Kress and Van Leeuwen, 2006). Google Vision API can capture the facial expression (angry, joy, surprise, sorrow) of each face and the location and size of each body in each image. Therefore, we extracted facial expressions and the number of faces in the images of each project to represent contact; for social distance, we extracted relative square ratio of largest human to calculate the long shot, medium shot, and close shot of images; we finally extracted color and composition variations to represent modality. We computed the standard deviation of color (i.e., warm/cool, saturation, brightness, and contrast) and the standard deviation of compositions among images of each project to gauge the color and composition variations.

The Textual Metafunction Representation

The textual metafunction enables the function with ideational and interpersonal meanings. The focus of the textual component will be on the analysis of lexical density and grammatical complexity. Halliday (1989) stated that written language becomes complex by being lexically dense since it packs a large number of lexical items into each clause, while spoken language is considered grammatically complex. He (1994)

[1] The modality here refers to a metafunction representation, which is different from a data modality (e.g., texts, images).

further developed the measurement of lexical density as "the proportion of content (lexical) words - basically nouns, full verbs, adjectives, and adverbs derived from adjectives - over the total number of words in a text". Thus, we extracted lexical words from the description texts and use the proportion of lexical words in texts to represent the textual meaning.

For images, the textual metafunction is integrated by the compositional relations of information value, framing, and salience of images (Kress and van Leeuwen, 2006). The information value shows the placement of elements within an image. They can be placed in the center or margin, left or right. Kress and van Leeuwen (2006) suggested that elements on the left side of the visual image are considered as something already known, while the right side presents new information. Therefore, we applied the location of the main body in each image to show the information value feature.

Framing refers to the visual devices used to connect or separate the content in an image. Color is the key feature of image framing because it connects or separates important objects within simple pictures (Norris, 2014). Salience refers to the ability of a visual participant to capture the viewer's attention. Salience is related to the visual weight of elements in a layout, and it is determined by a variety of features, such as the size of elements, tonal contrast, and color contrast. Therefore, we extracted relative features of the image colors to represent framing, and main object size and the color contrast to represent salience.

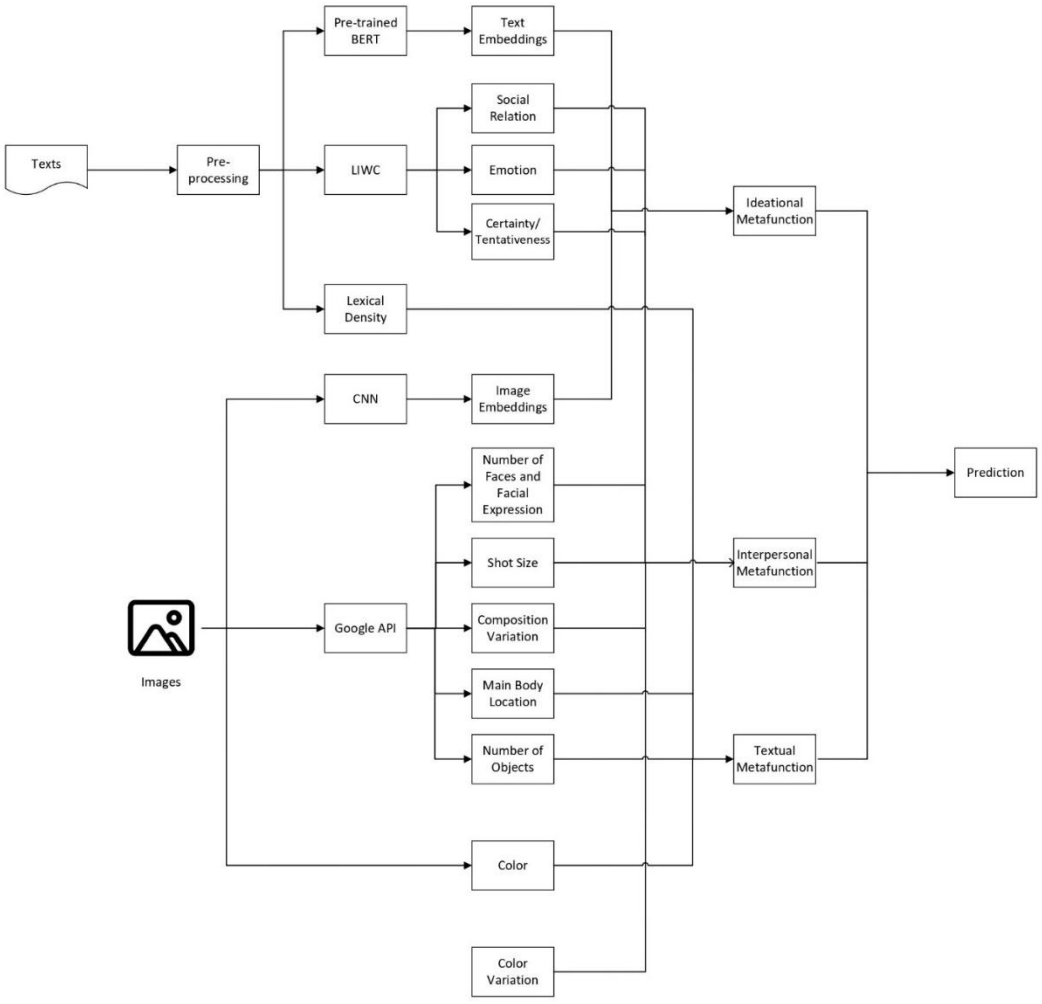

Figure 1. Outline of the Proposed Framework

Table 1 summarizes the proposed metafunction representations of multimodality, and Figure 1 outlines the proposed framework. First, we acquire multimodal data from the crowdfunding platform and then conduct a series of text preprocessing. Second, we extract the ideational, interpersonal, and textual representation

features from multimodal data using deep learning models (BERT, CNN), Google Vision API, and LIWC. Third, we build machine learning models to predict fundraising success on the three metafunction representation feature sets of multimodal data.

Data Modality	Ideational Metafunction	Interpersonal Metafunction	Textual Metafunction
Text	Experiential meanings, logical meanings (Textual elements)	Interaction and social relation	Lexical density
		Polarity	
		Modality	
Image	Visual elements	Contact	Information value
		Social distance	Framing
		Modality	Salience

Table 1. Metafunction Representations of Multimodality

Empirical Evaluation and Results

Data

We evaluated our proposed framework on a dataset collected from a famous reward-based crowdfunding platform, Kickstarter.com. We collected data about the projects that were launched between 2014 and 2019. Since the object types of images may vary a lot across different categories (e.g., there are more humans in the music category but fewer in the technology category), we selected the music and film & video categories, which are the top two categories of launched projects, for empirical study. Since our research focuses on the multimodality of projects, we kept the projects with multiple data modalities for evaluation. There are 16,924 completed projects in total, 12,104 of which succeeded. Next, we extracted the texts and images from the description page of each project. Besides the texts and images, we also extracted and standardized some meta features of each project, including goal, duration, description text length, number of images, and number of videos.

Experiments

To test the predictive utility of ideational, interpersonal, and textual metafunctions of multimodal data of crowdfunding, we conducted a series of experiments using five machine learning methods: logistic regression (LR), least absolute shrinkage and selection operator (LASSO), support vector machine (SVM), random forest (RF), and XGBoost. We included the meta features as the baseline model and then concatenated our proposed ideational features, interpersonal features, and textual features to the meta feature set to form the ideational, interpersonal, and textual models, respectively. We examined the prediction performance of models on each modality and their combination in the first three experiments. We compared the three metafunction feature sets, as well as the possible combinations of them, to examine possible interactions among the metafunctions in influencing the performance of prediction in the fourth experiment.

For every setting in the experiments, we estimated the prediction performance, in terms of the area under the curve (AUC) and Kolmogorov–Smirnov (KS) statistic (Massey Jr, 1951), through ten independent rounds of ten-fold cross-validations, resulting in 100 estimates for each metric.

Experiment 1: Ideational Metafunction of Multimodal Data

Table 2 summarizes the prediction performance (AUC and KS) of the five classification methods, based on the ideational metafunction of text modality, image modality, and their combination, respectively. Table 3 summarizes the results of the Friedman test and Dunn's pairwise post hoc test according to the AUC metric. Overall, the ideational metafunction of any data modality performed significantly better than the baseline model ($p < .001$), showing the value of ideational metafunction in crowdfunding success prediction. Comparing the predictive effectiveness of different data modalities, the metafunction of the multimodal

data generated the best prediction performance, outperforming the baseline, text, and image models (p < .001). This implies that synthesizing data modalities helps improve the predictive utility. Moreover, there was no statistically significant difference in predictive utility between texts and images on ideational metafunction. This shows that the ideational metafunction of each data modality was important and contributed almost equally to the predictive effectiveness.

Model	Metric	LR	LASSO	SVM	RF	XGB
Baseline	AUC	0.657 (0.016)	0.625 (0.015)	0.688 (0.015)	0.692 (0.013)	0.720 (0.015)
	KS	0.240 (0.026)	0.200 (0.025)	0.311 (0.027)	0.298 (0.021)	0.344 (0.024)
Text	AUC	0.792 (0.011)	0.701 (0.013)	0.800 (0.011)	0.758 (0.012)	0.806 (0.012)
	KS	0.449 (0.022)	0.313 (0.022)	0.464 (0.021)	0.386 (0.022)	0.471 (0.022)
Image	AUC	0.732 (0.017)	0.712 (0.015)	0.792 (0.013)	0.802 (0.013)	0.829 (0.012)
	KS	0.357 (0.026)	0.325 (0.025)	0.453 (0.024)	0.468 (0.025)	0.514 (0.022)
Text+ Image	AUC	0.817 (0.011)	0.749 (0.014)	0.847 (0.011)	0.822 (0.012)	0.862 (0.012)
	KS	0.489 (0.024)	0.381 (0.024)	0.553 (0.024)	0.501 (0.023)	0.571 (0.023)

Standard deviations are enclosed in parentheses.

Table 2. Prediction Performance of Ideational Metafunction

	Average Rank	Adjusted p-value of Pairwise Comparison		
		Baseline	Texts	Images
Baseline	4.00			
Text	2.61	<.001		
Image	2.39	<.001	.053	
Text+Image	1.00	<.001	<.001	<.001
Friedman χ^2: 1355.071 (p < .001)				

Table 3. Friedman Test and Post Hoc Dunn Test on Ideational Metafunction

Model	Metric	LR	LASSO	SVM	RF	XGB
Text	AUC	0.663 (0.015)	0.629 (0.014)	0.696 (0.015)	0.725 (0.012)	0.735 (0.013)
	KS	0.259 (0.026)	0.209 (0.023)	0.315 (0.025)	0.341 (0.022)	0.361 (0.023)
Image	AUC	0.687 (0.014)	0.649 (0.013)	0.708 (0.015)	0.734 (0.015)	0.750 (0.015)
	KS	0.285 (0.023)	0.227 (0.022)	0.329 (0.025)	0.345 (0.025)	0.374 (0.024)
Text+ Image	AUC	0.692 (0.014)	0.655 (0.014)	0.709 (0.015)	0.750 (0.013)	0.761 (0.014)
	KS	0.291 (0.023)	0.234 (0.023)	0.324 (0.025)	0.364 (0.024)	0.388(0.022)

Standard deviations are enclosed in parentheses.

Table 4. Prediction Performance of Interpersonal Metafunction

Experiment 2: Interpersonal Metafunction of Multimodal Data

In the second experiment, we compared the prediction performances (AUC and KS) of interpersonal features of each data modality and multimodality. As shown in Table 4 (results for the baseline model are the same as that in Experiment 1 and hence omitted), the interpersonal metafunction of any data modality showed better predictive capabilities over the baseline model (p < .001), demonstrating the value of the human interaction of texts and images on fundraising success prediction. Comparing interpersonal metafunction in different data modalities, images led to better prediction performance than texts, and the

multimodality (texts and images) outperformed single modalities (p < .001), showing the superior predictive capability of multimodality on interpersonal metafunction.

	Average Rank	Adjusted p-value of Pairwise Comparison		
		Baseline	Texts	Images
Baseline	3.81			
Text	3.02	<.001		
Image	2.97	<.001	<.001	
Text+Image	2.20	<.001	<.001	<.001
Friedman χ^2: 1193.808 (p < .001)				

Table 5. Friedman Test and Post Hoc Dunn Test on Interpersonal Metafunction

Experiment 3: Textual Metafunction of Multimodal Data

The prediction performance (Table 6) and Friedman test (Table 7) showed the predictive effectiveness of textual metafunction of the image modality and multimodality over the baseline model (p < .001), revealing the value of textual metafunction of image and multimodality in fundraising success prediction. Comparing the effectiveness of different data modalities, the textual metafunction of images significantly outperformed that of texts (p < .001), and the multimodality of metafunction performed significantly better than each single modality (texts or images) (p < .001). The results also show that combining modalities contributed to prediction performance improvement.

Model	Metric	LR	LASSO	SVM	RF	XGB
Text	AUC	0.666 (0.016)	0.626 (0.015)	0.570 (0.020)	0.719 (0.015)	0.731 (0.014)
	KS	0.263 (0.025)	0.206 (0.024)	0.123 (0.022)	0.334 (0.023)	0.355 (0.023)
Image	AUC	0.672 (0.015)	0.636 (0.014)	0.588 (0.041)	0.730 (0.015)	0.747 (0.015)
	KS	0.268 (0.024)	0.215 (0.023)	0.149 (0.046)	0.340 (0.027)	0.371 (0.025)
Text+ Image	AUC	0.677 (0.017)	0.641 (0.015)	0.599 (0.041)	0.745 (0.015)	0.757 (0.015)
	KS	0.275 (0.027)	0.220 (0.024)	0.174 (0.040)	0.361 (0.025)	0.384 (0.025)

Standard deviations are enclosed in parentheses.

Table 6. Prediction Performance of Textual Metafunction

	Average Rank	Adjusted p-value of pairwise comparison		
		Baseline	Texts	Images
Baseline	3.20			
Text	3.08	.702		
Image	2.25	<.001	<.001	
Text+Image	1.47	<.001	<.001	<.001
Friedman χ^2: 587.361 (p < .001)				

Table 7. Friedman Test and Post Hoc Dunn Test on Textual Metafunction

Experiment 4: Combination of Metafunctions of Multimodal Data

Table 8 summarizes the prediction performance (AUC and KS) of combinations of metafunctions on multimodal data (results for the three metafunctions individually are the same as those in Experiments 1 to 3 and hence omitted). The results show that the combination of ideational and interpersonal metafunctions significantly outperformed a single metafunction (p < .001), implying that adding an ideational or interpersonal metafunction feature set improves the prediction performance. There is no statistically

significant difference between ideational metafunction and the combination of ideational and textual metafunctions or between interpersonal metafunction and the combination of interpersonal and textual metafunctions, implying the weak predictive utility of adding textual metafunction on top of ideational or interpersonal metafunction. The combination of all three metafunctions did not always yield the best prediction performance. For three classification methods (LASSO, SVM, and RF), incorporating ideational and interpersonal metafunctions achieved the best performance. This implies potential conflicts or overfitting among metafunctions.

Model	Metric	LR	LASSO	SVM	RF	XGB
Ideational+ Interpersonal	AUC	0.832(0.011)	0.761(0.013)	0.843(0.012)	0.842(0.012)	0.884(0.011)
	KS	0.509(0.024)	0.389(0.022)	0.543(0.023)	0.522(0.024)	0.596(0.024)
Ideational+ Textual	AUC	0.830(0.011)	0.760(0.013)	0.774(0.012)	0.842(0.012)	0.885 (0.011)
	KS	0.507(0.022)	0.391(0.023)	0.446(0.022)	0.521(0.024)	0.596(0.023)
Interpersonal+ Textual	AUC	0.697(0.015)	0.660(0.015)	0.624(0.038)	0.750(0.014)	0.767 (0.014)
	KS	0.387(0.024)	0.335(0.023)	0.294(0.038)	0.364(0.025)	0.396(0.024)
All	AUC	0.832(0.012)	0.758(0.013)	0.772 (0.013)	0.842(0.012)	0.886(0.012)
	KS	0.511(0.024)	0.385(0.020)	0.444(0.021)	0.521 (0.021)	0.596(0.023)

Table 8. Prediction Performance of Metafunction Combinations

	Average Rank	p-value of pairwise comparison					
		Ideational	Interp-ersonal	Textual	Ideational+ Interperso-nal	Ideation-al+Textual	Interperson-al + Textual
Ideational	3.40						
Interpersonal	5.74	<.001					
Textual	6.78	<.001	<.001				
Ideational+ Interpersonal	1.97	<.001	<.001	<.001			
Ideational+ Textual	2.38	1.000	<.001	<.001	0.059		
Interpersonal+ Textual	5.48	<.001	1.000	<.001	<.001	<.001	
All	2.25	0.011	<.001	<.001	0.927	1.000	<.001
Friedman χ^2: 2476.413 (p < .001)							

Table 9. Friedman Test and Post Hoc Dunn Test on Metafunction Combinations

Discussion

Our experiments yield some interesting findings. First and foremost, discovering ideational, interpersonal, and textual metafunctions of multimodal data helps to improve the performance of fundraising success prediction. This finding reveals that experiential and logical meanings, social interactions, and compositions of project descriptions have predictive value. Second, comparing the predictive utilities of three metafunctions, our evaluation shows the better performance of ideational metafunction over interpersonal metafunction, which outperformed the textual metafunction, highlighting the important role of the experiential and logical meanings of a project.

In addition, our results show the superior predictive utility of multimodal data over a single modality, implying that increasing the diversity of data types is valuable for predicting fundraising success. With regard to the predictive utilities of modalities, it is interesting that interpersonal and textual metafunctions

conveyed by images were more valuable than those conveyed by texts in predicting fundraising success, while ideational metafunctions reflected by texts and images were almost equally effective. This finding implies that social connections or interactions and compositions are more effectively delivered by images. A possible reason may be that image is a better modality to interact with others, and visual elements are more communicative as social connection than textual elements.

Implications

Our study has implications for both research and practice. For research, to our knowledge, this is the first study rooted in metafunctions framework of languages (Halliday,1985), functional visual design (Kress and Van Leeuwen, 1996), and intersemiotic complementarity of languages and visual images (Royce, 1998) to discover metafunctions of multimodal data of crowdfunding projects. Our empirical study evaluated the effectiveness of ideational, interpersonal, and textual metafunctions in fundraising success prediction and demonstrated the predictive utility of any metafunctions and metafunction combinations. Moreover, our study showed that adding modalities of data can help to improve prediction performance. Third, we find interesting patterns among metafunctions conveyed by different data modalities. Specifically, the empirical results reveal a strong effect of interpersonal and textual metafunctions on image modality, but a weak effect of textual metafunction on text modality.

For practice, our work provides a framework for effective fundraising success prediction based on multimodal data. Effectively predicting the likelihood of success of a project is important for both entrepreneurs and investors. It informs entrepreneurs about their project potential, helping them adjust their campaigns, and at the same time, helps investors manage their funding risks and reduce opportunity costs.

Conclusion

As multimodal data become increasingly popular on crowdfunding platforms, entrepreneurs are willing to sell their ideas through a variety of modalities to make a good impression. There is a lack of research systematically exploring how functions of multimodal data take effect in predicting fundraising success. Our study identified metafunctions of multimodality in crowdfunding and demonstrated the predictive value of metafunctions and multimodality. Our empirical evaluation also showed how different metafunctions take effect on data modalities.

Our study has some limitations, which may be addressed in future research. First, while we have developed and evaluated the predictive capability of metafunctions of texts and images, other data modalities, such as audios and videos, remain to be evaluated. Second, while we have studied the predictive value of metafunctions in human-oriented project categories, the evaluation of their utilities needs to be extended to other categories, such as the technology category. Third, since we only examined the utilities of metafunctions on the Kickstarter platform, the generalizability of our findings needs to be validated on other reward-based platforms.

REFERENCES

Abbasi, A., and Chen, H. 2008. “CyberGate: A Design Framework and System for Text Analysis of Computer-Mediated Communication,” MIS Quarterly (32:4), pp. 811–837.

Alom, M. Z., Taha, T. M., Yakopcic, C., Westberg, S., Sidike, P., Nasrin, M. S., Hasan, M., Van Essen, B. C., Awwal, A. A. S., and Asari, V. K. 2019. “A State-of-the-Art Survey on Deep Learning Theory and Architectures,” Electronics (8:3), p. 292.

Cheng, C., Tan, F., Hou, X., and Wei, Z. 2019. “Success Prediction on Crowdfunding with Multimodal Deep Learning,” in Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, Macao, China: International Joint Conferences on Artificial Intelligence Organization, August, pp. 2158–2164.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. 2018. BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding.

Dong, W, Liao, S & Zhang, Z .2018. “Leveraging Financial Social Media Data for Corporate Fraud Detection,” Journal of Management Information Systems (35:2), pp. 461-487

Garrett, J. J. 2002. The Elements of User Experience: User-Centered Design for the Web / Jesse James Garrett., (1st ed.), Voices That Matter, Indianapolis, Ind.: New Riders.
Guijarro, A. 2010. "A Multimodal Analysis of The Tale of Peter Rabbit within the Interpersonal Metafunction," Atlantis (32), pp. 123–140.
Halliday, M. A. K. 1985. An Introduction to Functional Grammar, London: Edward Arnold.
Halliday, M. A. K. 1989. Spoken and Written Language, Oxford University Press, USA.
Halliday, M. A. K. 1994. An introduction to functional grammar, London: Edward Arnold.
Halliday, M.A.K. 2004. An Introduction to Functional Grammar, (3rd edition), London: Edward Arnold.
Hou, J.-R., Zhang, J., and Zhang, K. 2019. "Can Title Images Predict the Emotions and the Performance of Crowdfunding Projects," in Proceedings of the 52nd Hawaii International Conference on System Sciences.
Kress, G., & van Leeuwen, T. 1996. Reading Images: The Grammar of Visual Design, London: Routledge.
Kress, G., and Leeuwen, T. van. 2006. Reading Images: The Grammar of Visual Design, (2nd edition), London: Routledge.
Krizhevsky, A., Sutskever, I., and Hinton, G. E. 2012. "ImageNet Classification with Deep Convolutional Neural Networks," in Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1, NIPS'12, Red Hook, NY, USA: Curran Associates Inc., pp. 1097–1105.
Liu, N., Wan, L., Zhang, Y., Zhou, T., Huo, H., and Fang, T. 2018. "Exploiting Convolutional Neural Networks with Deeply Local Description for Remote Sensing Image Classification," IEEE Access (6), pp. 11215–11228.
Massey Jr, F.J., 1951. "The Kolmogorov-Smirnov test for goodness of fit," Journal of the American statistical Association, (46-253), pp.68-78.
Mollick, E. 2014. "The dynamics of crowdfunding: An exploratory study," Journal of Business Venturing (29-1), pp. 1–16.
Norris, S., and Maier, C. D. 2014. Interactions, Images and Texts: A Reader in Multimodality, Trends in Applied Linguistics, Hawthorne: Walter de Gruyter GmbH.
O'Halloran, K. 2008a. Mathematical Discourse: Language, Symbolism and Visual Images, A&C Black.
O'Halloran, K. L. 2008b. "Systemic Functional-Multimodal Discourse Analysis (SF-MDA): Constructing Ideational Meaning Using Language and Visual Imagery," Visual Communication (7:4), SAGE Publications, pp. 443–475.
Perez, B., Machado, S. R., Andrews, J. T. A., and Kourtellis, N. 2020. "I Call BS: Fraud Detection in Crowdfunding Campaigns," ArXiv:2006.16849 [Cs].
Royce, T. D. 1998. "Synergy on the Page: Exploring Intersemiotic Complementarity in Page-Based Multimodal Text. JASFL Occasional Papers (1:1), pp. 25-49.
Royce, T. D., Bowcher, W. (eds.). 2013. "Intersemiotic Complementarity: A Framework for Multimodal Discourse Analysis: Terry D. Royce," in New Directions in the Analysis of Multimodal Discourse (0 ed.), Routledge, pp. 71–117.
Simonyan, K., and Zisserman, A. 2014. "Very Deep Convolutional Networks for Large-Scale Image Recognition," ArXiv Preprint ArXiv, pp. 1409-1556.
Tausczik, Y. R., & Pennebaker, J. W. 2010. "The psychological meaning of words: LIWC and computerized text analysis methods," Journal of language and social psychology, (29-1), pp. 24-54.
Zappavigna, M., and Patrick, J. 2010. "Eliciting Tacit Knowledge about Requirement Analysis with a Grammar-Targeted Interview Method (GIM)," European Journal of Information Systems (19:1), pp. 49–59.
Zhou, M. (Jamie), Lu, B., Fan, W. (Patrick), and Wang, G. A. 2018. "Project Description and Crowdfunding Success: An Exploratory Study," Information Systems Frontiers (20:2), pp. 259–274.