Kiva.org: Exploring Relationships in Online Crowdfunding

Marc Valenti
Ines Alegre[1]
Xavier Tort[2]

[1]Department of Managerial Decision Sciences
IESE

[2]Department of Statistics and Operations Research
Universitat Politcnica de Catalunya

October 21, 2018

Findings

In Kiva, an online lending platform:

- Lending preferences (gender, country and sector) differ on lender characteristics (gender, country and profession).
- The descriptions written by the Partners can be attributed to them and visualized over time.
- The borrower's face expression determines the loan performance.

About Kiva: Definition and Data

Kiva is a non-profit organization. It is an online platform focused on micro-loans (with a median loan of 500$) that aims to connect borrowers and lenders. Most of this loans are for third world countries. 1.5M loans have been posted since their origin (2005), backed by more than 2.3M lenders.

Kiva has facilitated third parties access to their data by developing APIs and providing data snapshots. Thanks to this accessibility, its data has been used in 30 different research papers or thesis.

About Kiva: Agents

Kiva is the platform that connects **the three following agents**, agregates and transfers the capital. The three main agents in the platform:

- **Borrowers**, users that request loans. They can be charged interest rate fees from the **Partners**.
- **Partners**, mostly a local microfinance institution acting as a bridge between **Borrowers** and **Kiva**. Can charge interest rate, being reflected on **Borrower**'s side.
- **Lenders**, users that have the option of contributing with at least 25$ to the different available projects. **No interest rate is received** from the loans.

Objective and Hypothesis

The main objective of the thesis is to **contribute to the current research in online crowdfunding**. The following hypothesis are to be validated:

On the relationship between lenders and borrowers:

- **H1:** Lender characteristics (such as gender, country and profession) determine their lending preferences (as in gender, country and sector).

On the loan description:

- **H2:** Every partner has a template for their descriptions; being descriptions distinguishable across partners.
- **H3:** Partners may change their description template over time and copy other partners.

On borrowers' image:

- **H4:** Machine Learning can be used to extract expression labels on images.
- **H5:** The borrowers face expression on the image has an impact on the loan performance.

Chapter Hypothesis

On Lender-Borrower Relationship

H1: Lender characteristics (such as gender, country and profession) determine their lending preferences (as in gender, country and sector).

Litearture Review

- Galak, Small, and Stephen (2010) using Kiva: lenders prefer to give to those who are more like themselves (**gender, occupation, and first name initial**).
- Lin and Viswanathan (2013), using Prosper: transactions are more likely to occur between parties in the same geographical area
- Greenberg and Mollick (2015), using Kickstarter: projects founded by female founders are tended to be backed by more females (than projects not founded by females).

Summary

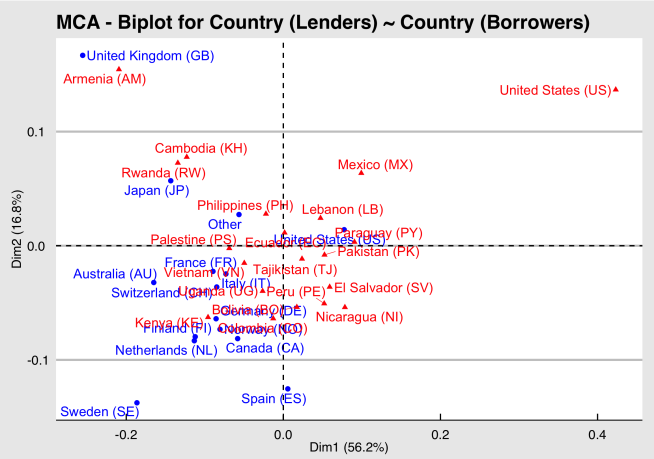
Continuous Data: Gender

Different models were suggested to find the relationship between the number of female borrowers and the number of female lenders (exhibiting a correlation of 0.2); and how this was affected by the sector the borrower was. None of them could be validated.

Categorical Data: Countries and Occupations

Used **Correspondence Analysis** and **Mosaic Plots** to explore similarities. Both χ^2 statistics reject the Null Hypothesis of independence, exhibiting then stronger deviations and therefore associations. This suggests that every Country has lending preferences. Finally, Correspondence Analysis has not been able to show *similar (based on history, region or race for example)* countries with *similar lending* behaviours.

Correspondence Analysis for Country



Chapter Hypothesis

On Loans Description:

- **H2:** Every partner has a template for their descriptions; being descriptions distinguishable across partners.
- **H3:** Partners may change their description template over time and copy other partners.

Text Processing

Bun T., 42, is married and lives in Kampong Cham Province with her six children. Her husband moved to work in Phnom Penh City as a construction worker with an income of US$8 per day. She has made up her mind that she wants to take a loan of US$700 to create a business of selling groceries at her house with the assistance of her children. If she succeeds in her business plan, she will find a job suitable for her husband to work back in his hometown.

bun t married lives kampong cham province six children husband moved work phnom penh city construction worker income us per day made mind wants take loan us create business selling groceries house assistance children succeeds business plan will find job suitable husband work back hometown

	business	children	city	groceries	house	husband	income	lives	loan	made	married	province	selling	wants	will	work	able	ahead	continues	grateful
120122	2	2	1	1	1	2	1	1	1	1	1	1	1	1	1	2	0	0	0	0
661165	2	0	1	0	0	0	1	1	0	0	0	0	0	0	1	0	1	1	1	1
251336	1	0	0	0	2	0	0	0	2	1	1	0	0	0	1	0	0	0	0	0
423290	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
421136	1	1	0	0	1	0	0	0	1	0	1	0	0	0	1	0	1	1	0	0

Dimensionality Reduction

Three different approaches are considered:[1]

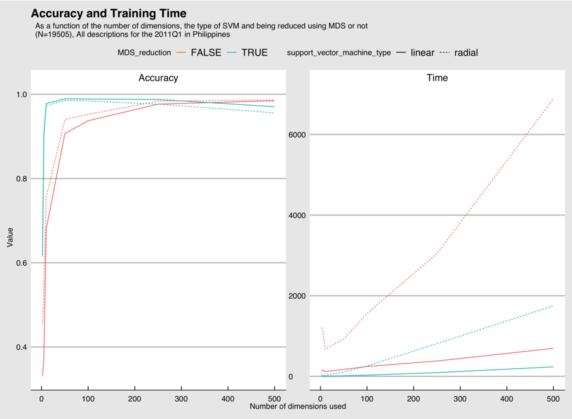
- Multidimensional Scaling using the Jaccard similarity
- Multidimensional Scaling using the Cosine similarity
- Multidimensional Scaling using Euclidean distance (Principal Component Analysis)

[1] See Figure 4.1.

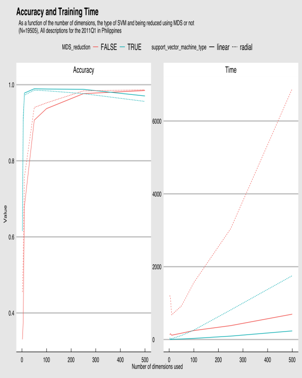
Evolution of Descriptions Over Time

- Evolution of Kenya Descriptions
- Evolution of Per Descriptions
- Evolution of Philippines Descriptions

Authorship Attribution



Authorship Attribution



Goal: For 2011Q1 in Philippines, attribute correctly the partner. [2]

- Implemented a Support Vector Machine classifier.
- 10-Fold Cross-Validation.
- 70% Train - 30% Test Split.
- Metrics: Accuracy on Test Set and Training Time on Training Set.

Outcome: $>$ 99% Accuracy achieved with 50 Dimensions from PCA and a Linear Kernel.

[2] Should be replicated for a later month.

Chapter Hypothesis

On Borrowers' Image:

- **H4:** Machine Learning can be used to extract expression labels on images.
- **H5:** The borrowers face expression on the image has an impact on the loan performance.

Litearture Review

Jenq, Pan, and Theseira (2011) , in Kiva: pictures and textual descriptions as determinants of individual charitable giving. Research Assistants rate each photograph on borrower's appearance, age, gender, perceived honesty, and skin color.

Data Collection

Google's Cloud Machine Learning Engine[3] and Microsoft's Azure Facial Recognition Software [4] are used to perform Image Recognition. They have been chosen because of their cutting edge technology and API access.

[3]https://cloud.google.com/vision/
[4]https://azure.microsoft.com/en-us/services/cognitive-services/face/

Data Collection: Image Retrieving + API Access

Process:

- Images were retrieved doing web-scrapping.
- Images are passed to the API as a POST method.

Only a subset of the data (N = 2696) is used. Different restrictions were imposed to reduce the variance.[5].

[5] Only a female borrower in the sector of Agriculture, Food or Retail, on the country of Philippines, working with partner 145 (*Negros Women for Tomorrow Foundation (NWTF)*), being the loan posted betweeen 2016-03-01 and 2016-04-01

Data Collection: Graphical Example

Figure: Image and Output from Loans 1033283 (left) and 1038440 (right)

Model Specification

Dependent variable: time to fund. The median time to fund is **90 hours.** Other **explanatory variables** included: `is_monday` (shorter time), `is_retail` (longer time) and `loan_amount` (longer time).
Two different sets suggested:

Manual_Scores

Manual selection of some scores obtained (higher interpretability).

- `happiness = G_joy+A_happiness-A_neutral`.
- `negative_others = .A_anger + A_disgust + G_sorrow + A_sadness + A_contempt + A_fear`.

Factor_Scores

Individual scores resulting by the Exploratory Factor Analysis. The first three dimensions are included:

- `FA_1` First Dimension Individual Scores of the **EDA**
- `FA_2` Second Dimension Individual Scores of the **EDA**
- `FA_3` Third Dimension Individual Scores of the **EDA**

Model Specification

Three different models are specified:

- Multiple Linear Regression Model
- Multiple Linear Regression Model with Dependent Variable transformed (log-linear Model)
- Logistic Model

Model Results

	Dependent variable:					
	(time_to_fund)	log(time_to_fund)	(time_to_fund)		log(time_to_fund)	(time_to_fund)
	OLS	*OLS*	*normal*	*OLS*	*OLS*	*normal*
	(1)	(2)	(3)	(4)	(5)	(6)
loan_amount	0.355***	0.002***	0.355***	0.355***	0.002***	0.355***
	(0.032)	(0.0002)	(0.032)	(0.032)	(0.0002)	(0.032)
is_monday	−29.795***	−0.316***	−29.795***	−29.514***	−0.314***	−29.514***
	(8.444)	(0.046)	(8.444)	(8.449)	(0.046)	(8.449)
is_retail	39.378***	0.180***	39.378***	39.358***	0.180***	39.358***
	(6.708)	(0.036)	(6.708)	(6.714)	(0.036)	(6.714)
happiness	−9.836***	−0.046***	−9.836***			
	(2.908)	(0.016)	(2.908)			
negative_others	−1.935	−0.095	−1.935			
	(20.190)	(0.109)	(20.190)			
FA_D1				−7.961***	−0.034**	−7.961***
				(2.504)	(0.014)	(2.504)
FA_D2				−4.470	−0.034*	−4.470
				(3.472)	(0.019)	(3.472)
FA_D3				0.391	−0.005	0.391
				(3.957)	(0.021)	(3.957)
Constant	24.226**	4.096***	24.226**	22.912**	4.086***	22.912**
	(10.442)	(0.056)	(10.442)	(10.421)	(0.056)	(10.421)
AIC	24055.7	4404.2	24055.7	24057.7	4405.5	24057.7
BIC	24094.5	4443	24094.5	24102.1	4449.8	24102.1
Observations	1,882	1,882	1,882	1,882	1,882	1,882
R^2	0.086	0.077		0.086	0.078	
Adjusted R^2	0.084	0.075		0.083	0.075	
Log Likelihood	−12,021.860		−12,021.860	−12,021.870		−12,021.870
Akaike Inf. Crit.			24,055.720			24,057.740
Residual Std. Error	144.018 (df = 1876)	0.778 (df = 1876)		144.057 (df = 1875)	0.778 (df = 1875)	
F Statistic	35.281*** (df = 5; 1876)	31.451*** (df = 5; 1876)		29.382*** (df = 6; 1875)	26.328*** (df = 6; 1875)	
Note:						*p<0.1; **p<0.05; ***p<0.01

Model Results

Best specification is the **Log-Linear Model**.[6].

As in model (2: Log-Linear Model and `Manual_Scores`), the coefficient for `happiness` is -0.046. The difference of a given image to be in the 75th percentile of happiness and the 25th percentile of happiness reduces the loan funding time by more than 10%.[7]

Model (5 Log-Linear Model and `Factor_Scores`) supports strongly with the previous argument while presenting an interesting Second Factor coefficient. [8]

[6] The validation of the model is shown in the original document

[7] Being the difference of the 75th and 25th percentiles of `happiness` $P_{75} - P_{25} = 2.19$. 10% is the result of (-0.046*2.19)

[8] The coefficient of the first dimension of EFA to be -0.034 (significant with $\alpha = 0.05$) and the coefficient of the second dimension of EFA to be -0.034 (not

Findings

In Kiva, an online lending platform:

- Lending preferences (gender, country and sector) differ on lender characteristics (gender, country and profession).
- The descriptions written by the Partners can be attributed to them and visualized over time.
- The borrower's face expression determines the loan performance.

Further Work

- Personalized feed drives to polarization.
- The resolution of the picture affects loan performance.
- The perceived age of the borrower in the picture affects loan performance.
- The sentiment analysis of the description affects loan performance.
- The interaction between the sentiment of the description and the face expression affects loan performance.

Thank You