"I Loan Because...": Understanding Motivations for Pro-Social Lending

Yang Liu
University of Michigan
Ann Arbor, MI, USA
owenliu@umich.com

Roy Chen
University of Michigan
Ann Arbor, MI, USA
roywchen@umich.edu

Yan Chen
University of Michigan
Ann Arbor, MI, USA
yanchen@umich.edu

Qiaozhu Mei
University of Michigan
Ann Arbor, MI, USA
qmei@umich.edu

Suzy Salib
University of Michigan
Ann Arbor, MI, USA
ssalib@umich.edu

ABSTRACT

As a new paradigm of online communities, microfinance sites such as Kiva.org have attracted much public attention. To understand lender motivations on Kiva, we classify the lenders' self-stated motivations into ten categories with human coders and machine learning based classifiers. We employ text classifiers using lexical features, along with social features based on lender activity information on Kiva, to predict the categories of lender motivation statements. Although the task appears to be much more challenging than traditional topic-based categorization, our classifiers can achieve high precision in most categories. Using the results of this classification along with Kiva teams information, we predict lending activity from lender motivation and team affiliations. Finally, we make design recommendations regarding Kiva practices which might increase pro-social lending.

Categories and Subject Descriptors

I.2.7 [**Natural Language Processing**]: text analysis

General Terms

Economics, Performance

Keywords

lending motivation, text classification, microfinance, pro-social lending, Kiva

1. INTRODUCTION

Understanding the motivation for pro-social behavior is the foundation for building a more realistic theoretical model of social preferences. Towards this end, experimental economists have used sophisticated experimental designs and econometric techniques to infer participants' motivations and social preferences in the lab [9]. While experimental data generated from the laboratory have yielded important insights into social preferences [14, 6], such results typically come from student subjects who engage in artificially-constructed games. This is necessary because social scientists rarely have the opportunity to record the self-articulated motivations of a large number of people as they engage in pro-social behavior in the real world. The growing popularity of microfinance provides a unique opportunity to explore this issue.

Globally, more than one billion people live in absolute poverty.[1] With few assets, most of these low-income households are excluded from the formal banking sector. To alleviate poverty, microfinance programs have emerged in many parts of the world to provide small loans and other financial services to the poor. Currently about 155 million households are served by microfinance programs, which help very poor households meet basic needs, improve household economic welfare, empower women, and promote entrepreneurship.[2]

Created in October 2005 as the first peer-to-peer microlending site, Kiva (kiva.org) matches citizen lenders with low-income entrepreneurs in developing countries.[3] Through Kiva's platform, anyone can make an interest-free loan of $25 or more to support an entrepreneur. As of August 2011, the total value of all loans made through Kiva was $233,051,800, 81% of which have been made to female entrepreneurs. When lenders register on the site, they have the option to fill in a field labeled "I loan because ...." About 100,000 lenders articulate these motivations on Kiva. Thus, in addition to its social impact on poverty alleviation, Kiva provides a unique data set with which we can study motivations for pro-social behavior.

This study classifies pro-social behavior outside the laboratory setting and uses the classified motivations and team affiliations to predict lending behavior, thus furthering our understanding of the motivations for such behavior. To do so, we draw on theories of social preferences and social identity to generate categories of motivation. We then train human coders to classify a randomly-selected sample of these statements. We use text classification techniques from machine learning to train classifiers on these hand-coded statements, which are then used to label the remaining state-

WSDM'12, February 8–12, 2012, Seattle, Washington, USA.

[1]In 2008, the World Bank revised its poverty cutoff to $1.25 per day at 2005 purchasing-power parity [36].

[2]MicroBanking Bulletin, Issue #19, December, 2009, pp. 49. Microfinance Information Exchange, Inc.

[3]More recently, through the Kiva City program, small business owners in the United States also can become beneficiaries of microlending on Kiva.

ments. We then use econometrics to predict lending behavior based on motivations and team information.

Text classification of user motivations is a novel, yet well-defined natural language processing task. However, it is more challenging than traditional topic-based classification tasks due to the relatively short text lengths of stated motivations and the subjectivity and subtleness of the motivations. Our work serves as a pioneering exploration of motivation classification. Our technique is applicable to other contexts where understanding user motivations is a concern.

Using the best-performing classifiers, the motivations of the 95k unlabeled statements are classified. Along with the information of Kiva teams, we predict lending activities from lender motivation and team affiliations. Finally, we make design recommendations regarding Kiva practices which might increase pro-social lending.

2. RELATED WORK

In general, our study is related to the literature of both computer science, especially text mining, and economics.

To the best of our knowledge, the classification of user motivations is not well covered in previous literature. The most closely-related work is the classification of user intent in search queries [28, 19]. An early classification scheme categorizes the intent of Web search queries into navigational, informational, and transactional [7]. More recent work identifies the missions and tasks in search sessions [22]. The goal of such work is to better understand user intent in order to improve the quality of results of search engines. Most such classification tasks are done based on the analysis of search engine logs rather than natural language processes. Among them, Daumè and Brill [12] induce web search intent via query reformulation which does not require click through data.

The subjectivity and subtleness of user motivations have distinguished our task from traditional topic-based text classification (e.g., politics vs. sports). This links our work to sentiment classification and opinion mining [34, 33]. Indeed, sentiment classification is widely considered to be a much more challenging task than topic-based classification. While the target categories of sentiment classification are usually simple and clear (e.g., positive vs. negative, like vs. dislike), the classification scheme of user motivations is usually not pre-defined - it largely depends on context, usually involves many more categories, and is usually distributed unevenly. As a result, motivation classification appears to be even more challenging than sentiment classification.

Furthermore, our work is also related to text classification of user-generated content and social media in general (e.g., [1, 30, 38]). For example, Agichtein et al.[1] have used classification methods to extract high-quality content from question/answer forums using both content and usage metadata features such as user relationships and usage statistics. Although our goal is fundamentally different from this body of work, the selection of techniques and features is certainly related.

In recent years, the study of microfinance in economics has grown substantially [4]. While they have historically offered low rates of default and good returns and growth [25], a major problem with microfinance that has received attention from economists is that of funding. As Bogan [5] points out, the demand for microfinance services far outstrips the supply. However, much of the recent economic literature on microfinance focuses on the demand (borrower) side, investigating the factors [23] and incentive mechanisms [17] affecting loan repayment. In comparison, we study the supply (lender) side. Specifically, we investigate the effects of lender motivations and team affiliations on lending behavior, neither of which, to our knowledge, has been studied in the economics literature.

3. DATA SOURCE AND KIVA STATISTICS

Using the data API provided by Kiva, we downloaded the motivation statements, team membership, and activity history of all lenders. All data used in this study will be made available.

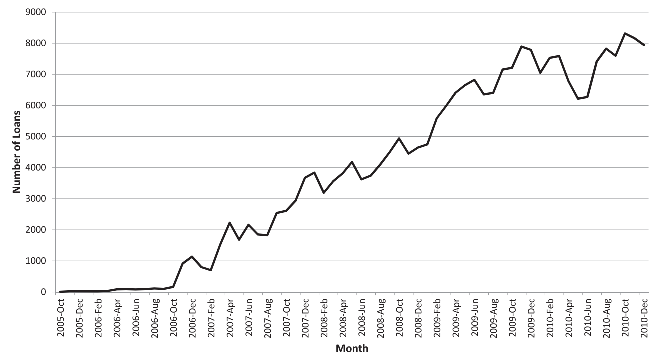

Figure 1: Number of loans funded through Kiva grows.

As of December 2010, Kiva has 660,183 registered lenders from 209 countries. The top five countries by the number of lenders are the United States, Canada, Australia, Great Britain, and Germany. Among all registered lenders, around two thirds of them have made at least one loan. Figure 1 shows the number of loans funded through Kiva each month from October 2005 to December 2010. We see that the number of loans has increased dramatically. As of December 2010, the number of loans made per month per lender is around 0.014.

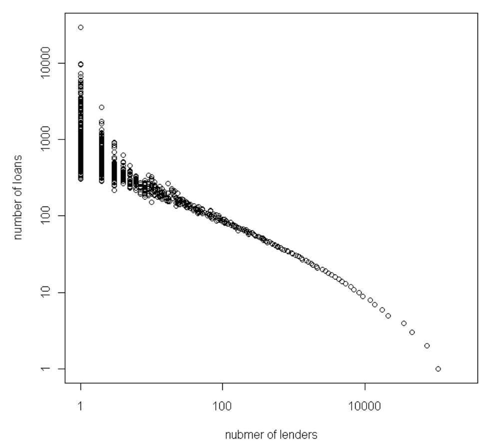

Figure 2: Distribution of lending frequency in log-log scale: few lenders made many loans; many lenders made few loans.

Figure 2 shows the distribution of the frequency of lending activity. The distribution follows a power law (characterized by a long tail). As mentioned above, one third of Kiva lenders have never made a loan. 106,511 (16.1%) lenders have only made one loan. A major problem within the Kiva community is that a large proportion of users are peripheral, contributing once or not at all, and only a few are core users who contribute frequently.

In August 2008, Kiva launched a new program supporting teams of lenders. This allowed users to join teams of other lenders, such as “Team Europe” or “Poverty2Prosperity.org - Poverty-Escape.”

The teams are displayed on a leaderboard ranked by the total amount loaned by its members. A lender can join more than one team. Figure 3 show that the number of teams that a Kiva user joined also follows a power law distribution. 85% of lenders have no team affiliation and 12% of lenders joined only one team. Very few people have joined multiple teams.

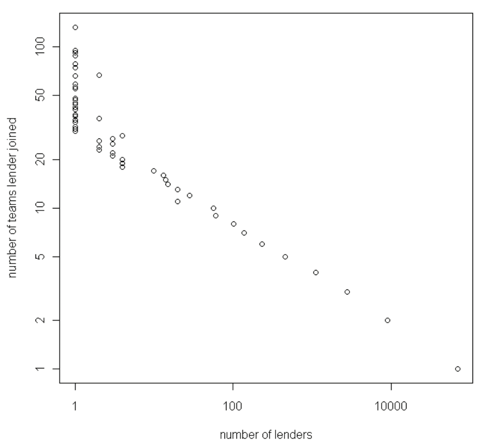

Figure 3: Distribution of lender joined teams in log-log scale: few lenders joined multiple teams; many lenders joined only one team.

At registration, Kiva allows a lender to write a short statement of her motivation for lending. It is interesting to notice that the lenders' motivations are usually closely related to the teams they join. To test whether lenders in same team have similar motivations in a quantitative way, we compute how similar the motivations of two users are, based on the the cosine similarity of the two statements. This enables us to evaluate how the motivational coherency (i.e., the average similarity of the motivation statements of every pair of team members) of each team. We then compare these intra-team similarities with a baseline computed as the average similarity of motivations among all users. The distribution of intra-team similarities is plotted in Figure 4. Among 1,185 Kiva teams having at least two members with a motivation statement, more than 1,000 of them have more coherent member motivations than the baseline (the horizontal line in Figure 4).

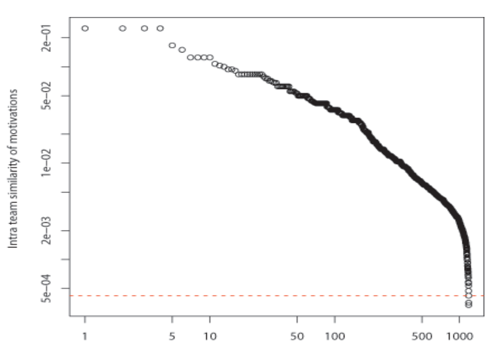

Figure 4: Distribution of intra-team similarity: most teams have more coherent motivations than the baseline.

This statistic suggests that a lender's motivation statement is likely to be correlated with her team membership, and could potentially predict her lending behavior. Is this necessarily true? We present a formal analysis to test whether lending activity is predictable from lender motivation and team affiliations.

4. METHOD

We combine research methods from text mining, experimental economics and econometrics. Our data analysis proceeds in four phases: theory-guided categorization, incentivized coding, text classification, regression analysis.

4.1 Theory-Guided Categorization

Based on theories of social preferences [14, 35] and social identity [3, 39] as well as our own understanding of the microfinance lending market, we develop an initial set of motivation categories for the individual "I loan because ..." statements. Two of the authors code a random sample of 200 individual statements independently and compare their coding assignments. They discuss any discrepancies until they agree. Based on these discussions, we revise the categories for each of the 200 statements into the following ten categories (with abbreviations in parentheses):

1. **General altruism** (Gnl. Altruism): e.g., "I believe in a global community."
2. **Group-specific altruism** (Grp. Altruism): e.g., "I want to help women succeed in business and in life."
3. **Empathy**: e.g., "I am disabled and I know what it's like to feel helpless."
4. **Reciprocity**: e.g., "I am very fortunate to have several people in my life to lend me a hand when I needed help. I hope that I can do the same for someone."
5. **Equality and social safety net** (Equity): e.g., "I want to help others who are less fortunate. Everyone deserves a fair chance."
6. **Social responsibility and social norms** (Norms): e.g., "I have the ability and I'm lucky enough to be able to."
7. **Effective development tool** (Tool): e.g., "I believe in change through bottom-up initiatives and sustainable business models."
8. **Personal satisfaction** (Satisfaction): e.g., "It makes my heart smile."
9. **Religious duty** (Religious): e.g., "I believe that sometimes God works thru people to answer prayers. What a privilege!"
10. **External reasons** (External): e.g., "It's for a community service project at my university."

4.2 Incentivized Coding Procedure

After determining our motivation categories, we have human coders code a randomly-selected sample of 5,250 statements, following the standard coding procedures in content analysis [26]. Each statement is coded by three independent coders.

To train the coders, we hold a one-hour in-person training session for 21 coders recruited from a database of University of Michigan students willing to participate in behavioral economics experiments. In this session, coders are introduced to microfinance, the Kiva web site, and the coding task at hand. We then describe the motivation categories in detail, and use a practice set of 50 random statements to train coders on the appropriate category for each of the different types of statements. Coders are encouraged to ask clarifying questions, which are answered in public. Coding instructions are available from the authors upon request. After the training session, each coder is asked to code 750 "I loan because. . ." statements and to log into a web interface to code the assigned statements remotely.

The computer interface used for the in-person training is the same as the remote interface used for the off-site coding by each of the human coders. Each human coder is assigned a unique login ID and a password to ensure the security of the coding sessions.

To encourage accurate coding, we employ an incentivized payment scheme. Recent experimental evidence indicates that coders are more responsive to classification criteria with incentivized payment based on correctness than with traditional piece-rate or flat-rate payment schemes [18]. Specifically, we pay coders a base rate of \$0.15 per statement, for a maximum possible base rate payment of \$112.50 if a coder finishes all 750 statements. To the base rate, we add the possibility of a bonus payment of up to \$20, based on the percentage of coded statements which agree with the authors' categorizations[4]. If a coder correctly codes 100% of the 750 statements assigned to her, she receives the full \$20 bonus. Otherwise, the bonus is calculated as the percentage of correct categorizations multiplied by \$20. This bonus is added to the base rate.

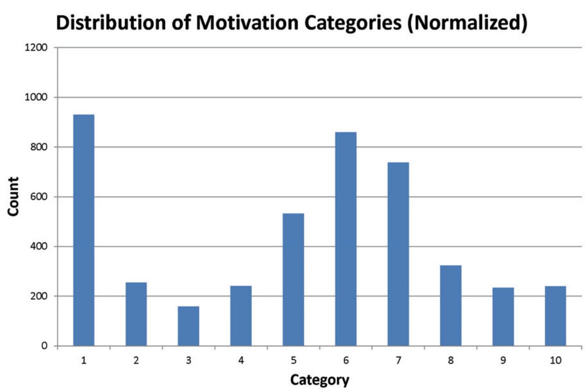

Figure 5: Distribution of Motivation Categories In Hand-Coded Sample (with normalization)

In the random sample of 5,250 statements, the distribution of motivations is presented in Figure 5. Note that this figure is normalized. Any statement can only contribute a total of 1 to the total count. If a statement is coded to two categories, for instance, then both of those categories receive a count of 0.5 for that statement.

We then examine the degree to which the coders agree with each other using interrater reliability. Interrater reliability is assessed with the intraclass correlation coefficient (ICC[3,3]),[5] which is a multi-rater generalization of Cohen's Kappa for the two-rater case.

	Rater Groups						
Category	1	2	3	4	5	6	7
1. Gnl. Altruism	0.50	0.47	0.44	0.53	0.44	0.54	0.52
2. Grp. Altruism	0.69	0.70	0.59	0.57	0.72	0.87	0.69
3. Empathy	0.61	0.76	0.64	0.64	0.48	0.82	0.55
4. Reciprocity	0.72	0.70	0.52	0.58	0.66	0.75	0.48
5. Equity	0.37	0.46	0.47	0.35	0.61	0.63	0.38
6. Norms	0.58	0.76	0.44	0.67	0.69	0.83	0.77
7. Tool	0.46	0.60	0.41	0.39	0.51	0.65	0.56
8. Satisfaction	0.73	0.69	0.52	0.54	0.74	0.77	0.59
9. Religious	0.89	0.87	0.72	0.85	0.87	0.93	0.86
10. External	0.67	0.68	0.31	0.57	0.65	0.74	0.50

Table 1: Intraclass Correlation Coefficients: ICC[3,3]

Table 1 reports the reliability statistics for the seven groups of raters. In general, values above 0.75 represent excellent reliability, values between 0.40 and 0.75 represent good reliability, and values below 0.40 represent poor reliability. We find that reliability varies across categories. Raters achieve good to excellent reliability in categories 1 (general altruism), 2 (group-specific altruism), 3 (empathy), 4 (reciprocity), 6 (social responsibility and social norms), 8 (personal satisfaction) and 9 (religious duty), and poor to good reliability among the remaining categories, indicating the challenge of classifying motivations.

4.3 Text Classification

The feasibility of human coding at a much larger scale is restricted by the availability of human and financial resources. We use the hand-coded motivations obtained through the above procedure to perform machine learning and train automatic text classifiers. We employ standard supervised and semi-supervised learning methods with different semantic, syntactic and social network features.

Since we do not obtain uniformly high inter-rater reliability across all ten categories, we conduct all experiments restricting our analysis to lender motivations for which all three coders agree on the motivation categories. Bear in mind that the purpose of our text classification is to accurately generate motivation categories for the 95k unlabeled lenders as one of the inputs for further regression analysis. The quality of the training data is critical to the the accuracy of our prediction. The restriction of the training set to unanimously coded data gives us higher confidence in the regression results. Of the original 5,250 motivations, 1,964 are unanimously coded in at least one category.

The distribution of unanimous motivations is presented in Figure 6. Note that this figure is normalized in the same way as Figure 5. From Figure 6 we see that the number of motivations in categories with low inter-rater reliability (e.g. category 1) drops more significantly than the number of motivations in categories with high inter-rater reliability (e.g. category 9).

We first process the statements stemmed by the Krovetz stemmer [27]. Stop words are not removed, as some stop words may be useful features for some classes (e.g., "I can."). We represent each document as a vector of features. By default, for each motivation statement, unigram, bigram, and parts-of-speech tags are extracted as features. These features are quantified either using a binary value or using a TF-IDF weight. No feature selection is applied.

[4]If the authors assign a statement to two or more categories, the human coder has to do the same, both in the number of categories and the specific assignments, to be considered correct.

[5]There are six main cases of intraclass correlation coefficients (ICC), distinguished by the numbers in parentheses following the letters ICC. The first number indicates the statistical model assumed. Case 3 assumes that judges are fixed and not drawn from a random population. The second number indicates the number of raters. More details on ICC computation can be found in [37].

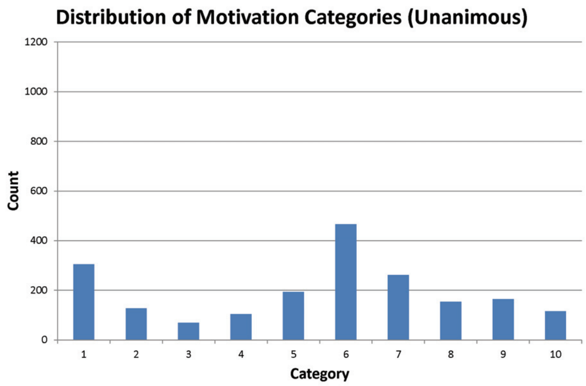

Figure 6: Distribution of Motivation Categories In Consentient Hand-Coded Sample (with normalization)

We then train a binary classifier for each category using Naïve Bayes [29], maximum entropy [31], and support vector machines (SVM) [11]. Note that we do not use multiclass classifiers directly because a statement can belong to more than one category. We adopt the the Naïve Bayes and maximum entropy classifiers from the Mallet package[6], and the SVM^{light} implementation[7] of the SVM classifiers. For the SVM classifiers, we explore different parameters, trading-off between training error and margin, and use two different kernels (i.e., a linear kernel and a RBF kernel).

Note that there are multiple interesting research issues beyond the application of a standard text classifier. First, since we are training a binary classifier for each category, the number of negative examples is far greater than the number of positive examples in the training data. This imbalance may result in the suboptimal performance of standard classification algorithms [20]. To address this issue, previous studies such as [10] have employed several approaches: cost-sensitive learning, minority-class oversampling and majority-class undersampling [40]. We use a model called "SVM-WEIGHT," which utilizes a cost-sensitive learning approach for SVM [2], implemented by libsvm. The basic idea of this algorithm is to penalize false negatives more heavily than false positives [32, 41]. We experiment with Tang et al (2009)'s approach to find the best value for the cost of a false negative.

Second, in our task there are far more unlabeled statements (i.e., 95k) than labeled statements. We therefore employ semi-supervised learning methods to utilize this unlabeled data in the classification. In particular, we use Transductive SVM [21] (also released in the SVMlight package), a typical transductive learning method, to bring unlabeled statements into the loop.

Finally, we intend to leverage the information of lender's social and lending activities in the classification tasks. The intuition here is that users' motivations are related to the teams they join and the number of loans they make. Thus the observation of such activity will in turn aid in the classification of motivations. We introduce specific features to the representation of a statement, such as the number of times a lender has loaned and the team(s) she has joined.

[6]McCallum, Andrew Kachites. "MALLET: A Machine Learning for Language Toolkit." http://mallet.cs.umass.edu. 2002.

[7]http://svmlight.joachims.org

Note that team membership may introduce many features as the number of teams is large. In our experiments, we first train a Naïve Bayes classifier using only team membership as the features, and then incorporate the output of this classifier as a meta-feature of the statement.

To train the text classifiers, we break the hand-coded data into training and test sets, and apply 5-fold cross-validation. The performance of each classifier is measured using a weighted F1 score (also referred to as the $F_{0.5}$ measure in some context [24]):

$$F_\beta = (1+\beta^2)\cdot\frac{precision \cdot recall}{\beta^2 \cdot precision + recall},$$

where β is set as 0.5. The reason we adopt the weighted F1 measure instead of the original F1 measure is that in the regression analysis, the precision of the motivation classification is a more important concern than the recall. For each category, we then apply the classifier with the highest performance on the hand-coded data to classify the rest of the 95k statements.

5. RESULTS: TEXT CLASSIFICATION

In this section, we present and discuss the performance of the text classifiers, assessed using the hand-coded statements. Note that the performances of classifiers with maximum $F_{0.5}$ measure are reported.

5.1 Standard Classifiers

We start with the performance of the standard classifiers, namely Naïve Bayes, maximum entropy, and SVM. Although we select the classifiers based on the $F_{0.5}$ measure which weights precision higher than recall, we also report the precision and recall of the classifiers.

Classifier	SVM	ME	NB
unigram presence	**72.62**	71.13	46.34
unigram tf-idf	**71.49***	55.86	23.82
bigram presence	65.13	**65.72**	36.45
bigram tf-idf	**65.84****	42.90	34.53
unigram+bigram presence	**73.17**	72.32	43.06
unigram+bigram tf-idf	**70.97*****	40.58	22.56
unigram+POS presence	**72.67**	70.62	46.45
unigram+POS tf-idf	**68.08*****	36.07	12.79

Table 2: Average $F_{0.5}$ measure of all classifiers with five-fold cross-validation, in percent. Boldface: best performance for a given row. Significant at the: * 10% level;** 5% level; *** 1% level. Note that SVM classifiers consistently outperform the other two families of classifiers

Table 2 summarizes the performance of each standard classification method averaging ten classes. Clearly, all classifiers perform significantly better than the random baseline $F_{0.5}$ measure of 10% (10% positive examples on average for ten categories). Among the three methods, we see that SVM classifiers with linear kernel consistently outperform the other two classifiers. When comparing the best performance of the SVM and maximum entropy classifiers, however, the difference between the best performers is not statistically significant (73.17 versus 72.32 with unweighted unigram and bigram features). Naïve Bayes performs significantly worse than the other two classifiers.

Table 3 presents the performance of the best SVM classifiers on each of the ten classes. The results suggest that some motivation categories lead to an easier classification task (for example, category 9: *religious duty*), while others present a greater challenge (e.g., category 5: *equality and social safety net* and category 7:

Category	$F_{0.5}$ measure (%)	Precision	Recall
Gnl. Altruism	67.87	72.15	55.20
Grp. Altruism	78.32	87.27	59.20
Empathy	74.81	82.24	55.97
Reciprocity	72.00	80.58	51.36
Equity	68.66	76.20	49.92
Norms	79.40	80.70	74.86
Tool	63.96	69.68	49.32
Satisfaction	75.51	84.54	53.45
Religious	88.21	95.60	67.63
External	62.91	74.25	39.70

Table 3: $F_{0.5}$ measure, precision, recall of classifers using unigram+bigram presence feature with 5-fold cross-validation

effective development tool). This is anticipated. In fact, the easy categories can usually be easily distinguished from others by keywords (e.g., for *religious duty*, "god", "prayers", etc.). The corresponding classification task is thus close to traditional topic-based classification. The other categories present much more subjectiveness and subtleness, where keywords and phrases do not have significant discriminative power. Another interesting perspective is to link the performance of automatic classification to the performance of human-coding. Indeed, the categories more "friendly" to the text classifiers are also associated with a higher inter-rater reliability (e.g., above 0.85 for *religious duty*), while the "classifier-resistant" categories are associated with a low inter-rater reliability (e.g., below 0.5 for *equality and social safety net*).

Furthermore, we find that, similar to the findings in sentiment classification [34], better performance is usually achieved when the features are not weighted (e.g., quantified with presence/absence only). This is because both sentiment classification and motivation classification documents are performed on short text (sentences and short statements) rather than rich documents (e.g., news articles). We anticipate that the presence of a feature conveys a strong signal in short documents, and that repeated appearance provides only marginal improvement. The combination of unigram and bigram features generally performs better than other combinations of features. Therefore, we adopt this combination in all following experiments.

5.2 Accommodating Imbalanced Data

Beyond the standard classifiers, we investigate the problem of handling imbalanced data (i.e. many more negative examples than positive examples in training). The performance of SVM-WEIGHT classifiers are reported in Table 4.

Compared to the standard classifiers reported in Table 3, we observe improved performance in four categories (1,5,6,9). However, none of the improvements is statistically significant. This can mostly be attributed to the high baselines achieved by the SVM classifiers. Since cost-sensitive learning moves the boundary towards the negative support vectors, a higher recall rate will be achieved at the expense of precision. The results suggest that over-weighting the minority class is effective in managing imbalanced datasets when precision is more important than recall.

5.3 Leveraging Unlabeled Data

Our second investigation is to leverage the unlabeled data in the classification - this is a plausible intuition behind semi-supervised learning. We incorporate an additional 19,000 randomly-selected (1/5 of the 95K available) unlabeled motivations into the training process using Transductive SVM. The results are summarized in Table 5. Surprisingly, we find that the use of transductive SVM results in a decrease of the $F_{0.5}$ measure in all ten categories.

Category	$F_{0.5}$ measure (%)	Precision	Recall
Gnl. Altruism	**68.71**	73.06	55.92
Grp. Altruism	78.32	87.28	59.20
Empathy	74.81	82.24	55.97
Reciprocity	72.00	80.58	51.36
Equity	**71.53**	94.13	36.58
Norms	**79.88**	80.92	76.47
Tool	63.96	69.68	49.32
Satisfaction	75.51	84.54	53.45
Religious	**88.44**	95.64	68.32
External	62.91	74.25	39.70

Table 4: The results of SVM-WEIGHT classifier with 5-fold cross-validation. Boldface: improvement over SVM.

Category	$F_{0.5}$ measure (%)	Precision	Recall
Gnl. Altruism	66.72	68.91	59.59
Grp. Altruism	71.35	70.00	77.76
Empathy	65.33	65.05	70.74
Reciprocity	64.81	64.24	67.93
Equity	59.66	61.71	56.07
Norms	73.71	71.64	83.50
Tool	61.75	63.81	55.74
Satisfaction	60.7	59.42	68.12
Religious	72.44	71.20	78.05
External	61.66	61.97	61.55

Table 5: The results of Transductive SVM with 5-fold cross-validation.

This seems to be inconsistent with the intuition of semi-supervised learning. With a more careful analysis, we observe that despite the decreased $F_{0.5}$ measure, the recall of all ten categories are largely improved. The improvement in recall for category 10 is more than 20 percentage points. With the help of the unlabeled data, the TSVM classifiers successfully generate more conservative boundaries towards the negative examples. In our precision-emphasized context, however, this doesn't lead to an increase of the $F_{0.5}$ measures. Also, the lexical properties of unanimously coded data are slightly different from that of the unlabeled data since the unanimously coded data have more salient features and are more easily classified. Besides transductive SVM, we also employ a number of classical graph-based semi-supervised learning methods (in particular, the methods proposed in Zhu et al (2003) [43] and Zhou et al (2004) [42]). A similar effect has been observed. Interestingly, similar patterns are reported in [15], which applied semi-supervised learning methods in the context of sentiment classification.

5.4 Leveraging Activity Features

Our next investigation goes beyond the text, incorporating signals from the activities of lenders. As presented in Section 4.3, features related to team membership and lending activities are incorporated into the SVM classifiers. The results are summarized in Tables 6 and 7. The involvement of activity features improve the performance of standard SVM classifiers in some categories. Interestingly, the only statistically significant improvement appears in one of the "hard" classes (category 5, *equality and social safety net*). None of the other categories present significant improvement. The results again imply a correlation between lenders' motivation categories and their lending and team joining activities.

5.5 Discussion

We have completed a systematic exploration of a new natural language processing task - the classification of user motivations. The subtleness and subjectiveness of user motivations have made

Category	$F_{0.5}$ measure (%)	Precision	Recall
Gnl. Altruism	**70.22**	77.89	51.42
Grp. Altruism	78.31	85.79	58.54
Empathy	68.61	80.89	49.18
Reciprocity	59.19	78.88	30.14
Equity	**71.35****	86.02	43.75
Norms	78.36	80.17	71.98
Tool	61.37	70.43	41.29
Satisfaction	69.24	79.02	46.86
Religious	88.11	94.75	69.70
External	**69.01**	82.45	43.22

Table 6: The classification results with unigram+bigram pres+team feature with 5-fold cross-validation. Boldface: improvement over SVM. Significant at the: ** 5% level

Category	$F_{0.5}$ measure (%)	Precision	Recall
Gnl. Altruism	64.44	63.41	69.54
Grp. Altruism	**81.11**	88.96	61.47
Empathy	73.69	83.91	51.83
Reciprocity	57.38	57.98	56.37
Equity	**73.25**	87.37	47.73
Norms	79.19	81.47	71.63
Tool	**65.08**	74.66	43.19
Satisfaction	72.89	84.28	47.82
Religious	87.49	94.55	67.51
External	**71.98**	84.12	46.55

Table 7: The classification results with unigram+bigram pres+loan times feature with 5-fold cross-validation. Boldface: improvement over SVM.

the problem much more challenging than common text classification tasks that are based on topics (e.g., Reuters, 20 newsgroups, political vs. sports). This difficulty is not only observed in the classification results, but also in the human coding results (i.e., low inter-rater reliability). Indeed, the closer the motivation category is to a topic (e.g., "religious duty"), the more discriminative power keyword features have, and the better text classifiers perform. Despite this challenge, a standard SVM classifier with unigram and bigram features still achieves reasonable performance over most of the categories.

The most closely related classification task is perhaps sentiment classification, which also has subtle and subjective classes. Indeed, many observations similar to ours can be found in the literature of sentiment classification. The major challenge here is the lack of a natural definition of categories in user motivation (e.g., positive vs. negative). Therefore, there is little established domain knowledge that can be utilized. Keyword matching using a simple list/lexicon of sentimental words can perform reasonably well (with an accuracy up to 69% reported in [34]). Unfortunately, such a resource does not exist in the context of motivation classification. Another additional challenge comes from the imbalanced distribution of classes.

Our exploration also provides useful insights into the future development of motivation classification. Although neither the treatment of imbalanced data nor the use of unlabeled data have brought significant improvement in our precision-driven context, they may help significantly in other scenarios where recall is more of a concern. On the other hand, the incorporation of user activity information has brought considerable improvement even though it is explored in a rather simple way. This suggests a promising direction for inferring a user's motivation from her behavior rather than from a motivation statement, especially in a context where rich social activity data is available. This also strengthens our hypothesis that the lending behaviors of Kiva users are predictable from their motivations.

Please note that in the regression analysis that follows, we do not use the classification results with activity features involved. This is because such information overlaps with some of our dependent/independent variables (e.g., lending amount and team membership). We thus classify the unlabeled statements with the best performer in Table 3 and use the results in the regression analysis.

6. RESULTS: LENDING BEHAVIOR

In this section, we first report regression analysis relating motivation categories and team affiliations to lending behavior. We then discuss design implications based on our regression results.

6.1 Regression Analysis

To evaluate how lender motivations affect their lending behavior, we run several ordinary least squares (OLS) regressions. The dependent variables are either (a) the average number of loans that a lender gives per month, or (b) the amount that a lender lends per month.

Table 8 reports four OLS regressions investigating factors affecting the average number of loans a user makes per month, i.e., loan frequency. The independent variables include lender motivations and their team affiliation information. Column (1) reports the first specification where only lender motivations are included as independent variables. Columns (2) - (4) report three more regressions, where we control for the number of teams a user has joined. We do this in three different ways. In column (2), we simply control for whether or not the user has joined at least one team. In column (3), we assume that the number of teams a user has joined affects behavior linearly, and therefore include the number of teams the user has joined as a regressor. Finally, in column (4), we again control for the number of teams a user has joined, but nonlinearly (i.e. we include a dummy variable for whether the user has joined 1 team, 2 teams, etc.).

Table 8 shows robust motivation and team activity effects on lending frequency, as the significance and direction of these effects do not change between specifications. Specifically, categories 1 (general altruism), 2 (group-specific altruism), and 10 (external reasons) negatively affect lending frequency. A lender motivated by general or group-specific altruism on average makes 0.11 fewer loans per month than others. The general altruism category, e.g., "I care," can be viewed as a catch-all category, both by Kiva lenders and by our coders. Lenders in this category gave non-specific statements about why they lend, perhaps indicating a lesser degree of motivation to lend than users who gave very specific reasons for lending. Users in the group-specific altruism category, on the other hand, may be more selective regarding the projects they lent to, denoted by their naming specific groups to which they wished to lend. Finally, lenders with external reasons to lend, such as fulfilling a required school project or as a recipient of a Kiva gift card, make 0.16 fewer loans per month than others. These lenders might be less intrinsically motivated compared to others on Kiva.

By contrast, categories 7 (effective development tool) and 9 (religious duty) both positively affect lending frequency. A lender who sees Kiva as an effective development tool makes 0.17 more loans per month than others. Their motivation statements indicate that they believe Kiva to be a better way to help the poor than through other means. While other Kiva lenders might also utilize other methods of helping, such as direct charitable donation, lenders in this category might be more likely to use Kiva than other methods. Of all motivation categories, category 9 (religious duty) has by far the largest effect on lending frequency. A lender motivated

Table 8: OLS Regressions of Motivations and Team Activity on Lending Frequency

Dependent Variable: Number of Loans Per Month				
	(1)	(2)	(3)	(4)
Gnl. Altruism	-0.12***	-0.12***	-0.12***	-0.11***
	(0.044)	(0.043)	(0.043)	(0.043)
Grp. Altruism	-0.14**	-0.13**	-0.12**	-0.11**
	(0.057)	(0.057)	(0.056)	(0.056)
Empathy	-0.10	-0.08	-0.07	-0.06
	(0.086)	(0.086)	(0.085)	(0.085)
Reciprocity	-0.08	-0.05	-0.05	-0.05
	(0.071)	(0.070)	(0.070)	(0.070)
Equity	0.02	0.01	0.01	0.01
	(0.054)	(0.054)	(0.054)	(0.053)
Norms	-0.00	-0.02	-0.01	-0.01
	(0.034)	(0.033)	(0.033)	(0.033)
Tool	0.19***	0.17***	0.16***	0.17***
	(0.037)	(0.037)	(0.037)	(0.037)
Satisfaction	-0.07	-0.05	-0.06	-0.05
	(0.059)	(0.058)	(0.058)	(0.058)
Religious	0.27***	0.24***	0.25***	0.25***
	(0.061)	(0.061)	(0.060)	(0.060)
External	-0.26***	-0.18**	-0.18**	-0.16**
	(0.081)	(0.080)	(0.080)	(0.079)
≥1 Team		0.78***		
		(0.025)		
# Teams			0.42***	
			(0.008)	
1 Team				0.53***
				(0.027)
2 Teams				0.82***
				(0.053)
3 Teams				1.09***
				(0.087)
4 Teams				1.71***
				(0.134)
5 Teams				2.60***
				(0.204)
6 Teams				4.09***
				(0.279)
7 Teams				4.71***
				(0.346)
8 Teams				1.43***
				(0.416)
≥9 Teams				11.51***
				(0.234)
Constant	0.64***	0.43***	0.46***	0.43***
	(0.015)	(0.017)	(0.015)	(0.016)
# Obs.	100240	100240	100240	100240
R^2	0.001	0.011	0.026	0.036

Notes: Standard errors in parentheses
Significant at the: *** 1%, ** 5%, or * 10% level

by religious duty makes 0.25 more loans per month than others. Social identity research finds that a salient group identity increases contribution to public goods [13]. We argue that religious identities are made salient on Kiva through its lending teams program. Since its inception in August 2008, the top two lending teams (in total amount loaned) have consistently been the Atheists (first place) and the Kiva Christians (second place), each featured prominently on the team leaderboards. Such identity-based team competition should motivate the team members to lend more.

When controlling for team affiliation (column 2), we find that a lender belonging to any team(s) makes 0.78 more loans per month than those without any team affiliation. Furthermore, assuming linearity (column 3), belonging to an additional team is associated with 0.42 more loans per month. Lastly, column (4) separately estimates the effects of belonging to different number of teams without assuming linearity. Overall, the positive effect of team affiliation on lending frequency is consistent with the predictions of social identity theory. Ethnographic studies of Kiva teams reveal that teams communicate through the Kiva message board [16], set specific goals with deadlines, and coordinate team activities by singling out specific loans to the team with the goal of raising 100% of the money for each loan ("loan-a-thon"). Although we are not aware of any systematic investigation of the effects of teams on lending, we conjecture that the ability of teams to communicate, coordinate and compete might contribute to the increased lending activity of team members.

In addition to lending frequency, we are also interested in the effects of motivation categories and team affiliation on the amount lent. However, to protect lender privacy, individual loan amount is not available through Kiva data API. Therefore, for this analysis, we employ a proxy variable for the amount lent. We know the list of projects that each lender lends to, as well as the total amount lent to each project. We therefore assume that each lender to a project lends an equal amount. Once we apply this assumption to all projects, we have a proxy for the total amount lent by each user.

Table 9 presents four OLS regressions using the proxy lending amount as the dependent variable. Independent variables in each regression are the same as those in Table 8. While the significance and direction of motivation categories and team effects remain the same as those in Table 8, it is informative to highlight the size of some of these effects. Specifically, a lender motivated by general or group-specific altruism lends $6 less per month than others, while those motivated by external reasons lend approximately $7 less. By contrast, a lender who sees Kiva as an effective development tool lends $5 more per month, while one motivated by religious duty lends $9 more. Again, when controlling for team affiliation (column 2), we find that a lender belonging to any team(s) lends $31 more per month than those without any team affiliation, while each additional team joined is associated with $16 more lent per month. Overall, the effects of motivation categories and team affiliation on amount lent is consistent with those on lending frequency.

Even though team affiliation is positively correlated with both the lending frequency and lending amounts, we do not rule out the possibility of a selection issue, in that lenders who join teams are perhaps inclined to lend more in the first place. We are collecting additional data in ongoing work to account for this possibility.

It is also important to note that the reliability of these regression results depends on the effectiveness of the classifier, which in turn depends on the quality of the human coded sample. Though we attempt to minimize the disagreement between the human coders through incentivized coding procedure, these regression results are still affected by these disagreements.

Table 9: OLS Regressions of Motivations and Team Activity on Lending Amount

Dependent Variable: Average Lent Per Month (Proxy)				
	(1)	(2)	(3)	(4)
Gnl. Altruism	-5.72***	-5.93***	-5.81***	-5.66***
	(1.759)	(1.750)	(1.739)	(1.731)
Grp. Altruism	-6.28***	-5.81**	-5.70**	-5.26**
	(2.296)	(2.286)	(2.270)	(2.260)
Empathy	-4.56	-3.78	-3.52	-3.04
	(3.483)	(3.467)	(3.444)	(3.428)
Reciprocity	-3.90	-2.79	-2.80	-2.76
	(2.861)	(2.848)	(2.829)	(2.816)
Equity	0.29	-0.00	-0.12	0.01
	(2.196)	(2.186)	(2.171)	(2.161)
Norms	-1.34	-2.05	-1.78	-1.62
	(1.360)	(1.354)	(1.344)	(1.338)
Tool	6.03***	5.07***	4.93***	5.08***
	(1.503)	(1.496)	(1.486)	(1.479)
Satisfaction	-3.43	-2.65	-3.04	-2.89
	(2.373)	(2.362)	(2.346)	(2.336)
Religious	10.13***	8.76***	9.53***	9.14***
	(2.469)	(2.458)	(2.441)	(2.430)
External	-10.73***	-7.57**	-7.33**	-6.71**
	(3.265)	(3.251)	(3.229)	(3.215)
≥1 Team		30.58***		
		(0.996)		
# Teams			15.93***	
			(0.331)	
1 Team				20.94***
				(1.111)
2 Teams				32.12***
				(2.158)
3 Teams				44.05***
				(3.539)
4 Teams				64.43***
				(5.408)
5 Teams				99.57***
				(8.262)
6 Teams				154.70***
				(11.301)
7 Teams				170.42***
				(14.022)
8 Teams				54.79***
				(16.830)
≥9 Teams				439.25***
				(9.475)
Constant	26.33***	18.11***	19.54***	17.97***
	(0.615)	(0.668)	(0.624)	(0.660)
# Obs.	100240	100240	100240	100240
R^2	0.001	0.010	0.023	0.032

Notes: Standard errors in parentheses
Significant at the: *** 1%, ** 5%, or * 10% level

6.2 Design Implications

Our regression analysis of lender motivation and team affiliation on lending behavior suggests that some Kiva practices can be improved to increase participation and commitment.

We find that lenders motivated by external reasons, such as those receiving a Kiva gift card from a friend, make 0.16 fewer loans and lend $7 less per month than others. This suggests that recruiting newcomers through gift cards or social networks[8] might not be sufficient to make newcomers commit to the Kiva cause. For instance, the Kiva gift cards act as another way for existing users to lend on Kiva rather than a way to bring in new active members. Lenders recruited through such channels are likely to become peripheral users. As it stands, Kiva has a large number of peripheral users. Recall one-third of Kiva users have never made a loan and 16% have only made one loan. Socializing newcomers and motivating peripheral participants to become active contributors is a core issue facing Kiva.

Our finding that lenders belonging to any team(s) make 0.78 more loans and lend $31 more per month than those without team affiliations, combined with ethnographic studies of Kiva teams, suggest that successful teams (measured by total amount loaned) might be an effective mechanism to socialize newcomers and to motivate peripheral participants.

After a new lender joins Kiva through one of its existing channels, Kiva should encourage them to join an active and successful team. Team recommendation could be based on the similarity between the newcomer motivation and existing team member motivations.

7. CONCLUSION

Understanding user motivations in online communities helps the analysis and modeling of user behavior. In this paper, we study the novel problem of classifying user motivation statements from Kiva, a well-known online microfinance community. An incentivized coding procedure is employed to generate human-labeled datasets for this text classification task. Despite the specific challenges of this task, we find that SVM-based classifiers using unigram and bigram features work reasonably well. However, the use of primitive community-based features does not significantly improve classification performance.

It is clear that some categories of user motivations are more difficult to identify than others. In our future work, we will pursue deeper linguistic features, both syntactic and semantic, to enhance the SVM classifiers. In addition, a richer set of social-behavioral features will be explored to further improve the classification task.

We also examine which categories are associated with changes in lending behavior and found both categories that increased and decreased lending. These indicate that Kiva should reconsider policies that will create peripheral lenders and focus on those that encourage users to become core contributors. While Kiva gift cards and their more recent “Help Kiva branch out” campaign might not be very effective, further development of the Kiva lending teams may have a beneficial effect on lending. To further study this idea, the next step in this research is for us to correctly control for selection bias in users’ joining of Kiva teams, thus giving us insight as to whether the act of joining teams increases lending.

[8] A more recent example is the “Help Kiva branch out" campaign from August 1, 2011 to August 13, 2011, when Kiva lenders are encouraged to invite their friends to join Kiva. Kiva provides free trial loans to the first 4,000 new users who make a loan through this campaign.

8. ACKNOWLEDGMENTS

The financial support from the National Science Foundation through grants no. SES-0720943, IIS-1054199, BCS-1111019 and a Google Research Award is gratefully acknowledged. We thank Xuan Liang for her excellent research assistance in the early phase of the project.

9. REFERENCES

[1] E. Agichtein, C. Castillo, D. Donato, A. Gionis, and G. Mishne. Finding high-quality content in social media. In *Proceedings of the international conference on Web search and web data mining*, pages 183–194. ACM, 2008.

[2] R. Akbani, S. Kwek, and N. Japkowicz. Applying support vector machines to imbalanced datasets. *Machine Learning: ECML 2004*, pages 39–50, 2004.

[3] G. A. Akerlof and R. E. Kranton. *Identity Economics: How Our Identities Shape Our Work, Wages, and Well-Being*. Princeton University Press, Princeton, New Jersey, 2010.

[4] B. Armendáriz and J. Morduch. *The Economics of Microfinance*. The MIT Press, Cambridge, Massachusetts, second edition, 2010.

[5] V. L. Bogan. Capital structure and sustainability: An empirical study of microfinance institutions. *Review of Economics and Statistics*, 2011. forthcoming.

[6] G. E. Bolton and A. Ockenfels. Erc: A theory of equity, reciprocity, and competition. *American Economic Review*, 90(1):166–193, March 2000.

[7] A. Broder. A taxonomy of web search. In *ACM Sigir forum*, volume 36, pages 3–10. ACM, 2002.

[8] C.-C. Chang and C.-J. Lin. *LIBSVM: a library for support vector machines*, 2001. Software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`.

[9] G. Charness and M. Rabin. Understanding social preferences with simple tests. *Quarterly Journal of Economics*, 117(3):817–869, August 2002.

[10] N. V. Chawla, N. Japkowicz, and A. Kotcz. Editorial: special issue on learning from imbalanced data sets. *SIGKDD explorations*, 6(1):1, 2004.

[11] C. Cortes and V. Vapnik. Support-vector networks. *Machine learning*, 20(3):273–297, 1995.

[12] H. Daumé III and E. Brill. Web search intent induction via automatic query reformulation. In *Proceedings of HLT-NAACL 2004: Short Papers*, pages 49–52. Association for Computational Linguistics, 2004.

[13] C. C. Eckel and P. J. Grossman. Managing diversity by creating team identity. *Journal of Economic Behavior & Organization*, 58(3):371–392, November 2005.

[14] E. Fehr and K. M. Schmidt. The theory of fairness, competition, and cooperation. *Quarterly Journal of Economics*, 114(3):817–868, August 1999.

[15] A. B. Goldberg and X. Zhu. Seeing stars when there aren't many stars: graph-based semi-supervised learning for sentiment categorization. In *Proceedings of the First Workshop on Graph Based Methods for Natural Language Processing*, TextGraphs-1, pages 45–52, Stroudsburg, PA, USA, 2006. Association for Computational Linguistics.

[16] S. E. Hartley. Kiva.org: Crowd-sourced microfinance and cooperation in group lending. Harvard University Working Paper, 2010.

[17] N. Hermes and R. Lensink. The empirics of microfinance: what do we know? *The Economic Journal*, 117(517):F1–F10, 2007.

[18] D. Houser and E. Xiao. Classification of natural language messages using a coordination game. *Experimental Economics*, 14:1 – 14, 2011.

[19] B. Jansen, D. Booth, and A. Spink. Determining the informational, navigational, and transactional intent of web queries. *Information processing management*, 44(3):1251, 2008.

[20] N. Japkowicz and S. Stephen. The class imbalance problem: A systematic study. *Intelligent Data Analysis*, 6(5):429–449, 2002.

[21] T. Joachims. Transductive inference for text classification using support vector machines. In *Proceedings of ICML*, pages 200–209. ACM, 1999.

[22] R. Jones and K. Klinkner. Beyond the session timeout: automatic hierarchical segmentation of search topics in query logs. In *Proceeding of the 17th ACM conference on Information and knowledge management*, pages 699–708. ACM, 2008.

[23] D. Karlan. Using experimental economics to measure social capital and predict financial decision. *American Economic Review*, 95(5):1688–1699, December 2005.

[24] R. Klinger and C. M. Friedrich. User's choice of precision and recall in named entity recognition. In *Proceedings of the International Conference RANLP-2009*, pages 192–196, Borovets, Bulgaria, September 2009. Association for Computational Linguistics.

[25] N. A. Krauss and I. Walter. Can microfinance reduce portfolio volatility? *Economic Development and Cultural Change*, 58(1):85–110, October 2009.

[26] K. Krippendorff. *Content analysis: An introduction to its methodology*. Sage Publications, Thousand Oaks, CA, 2nd edition, 2003.

[27] R. Krovetz. Viewing morphology as an inference process. In *Proceedings of the 16th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 191–202. ACM, 1993.

[28] X. Li, Y. Wang, and A. Acero. Learning query intent from regularized click graphs. In *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*, pages 339–346. ACM, 2008.

[29] M. Maron. Automatic indexing: an experimental inquiry. *Journal of the ACM (JACM)*, 8(3):404–417, 1961.

[30] G. Mishne. Experiments with mood classification in blog posts. In *Proceedings of ACM SIGIR 2005 Workshop on Stylistic Analysis of Text for Information Access*, page 19. Citeseer, 2005.

[31] K. Nigam, J. Lafferty, and A. McCallum. Using maximum entropy for text classification. In *IJCAI-99 workshop on machine learning for information filtering*, volume 1, pages 61–67. Citeseer, 1999.

[32] E. Osuna, R. Freund, and F. Girosi. Training support vector machines: an application to face detection. In *cvpr*, page 130. Published by the IEEE Computer Society, 1997.

[33] B. Pang and L. Lee. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1-2):1–135, 2008.

[34] B. Pang, L. Lee, and S. Vaithyanathan. Thumbs up?: sentiment classification using machine learning techniques. In *Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10*, pages 79–86, 2002.

[35] M. Rabin. Incorporating fairness into game theory and economics. *American Economic Review*, 83(5):1281–1302, December 1993.

[36] M. Ravallion, S. Chen, and P. Sangraula. Dollar a day revisited. *World Bank Economic Review*, 23(2):163 – 184, 2009.

[37] P. E. Shrout and J. L. Fleiss. Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, 86(2):420–428, 1979.

[38] B. Sriram, D. Fuhry, E. Demir, H. Ferhatosmanoglu, and M. Demirbas. Short text classification in twitter to improve information filtering. In *Proceeding of the 33rd international ACM SIGIR conference on research and development in information retrieval*, pages 841–842. ACM, 2010.

[39] H. Tajfel and J. Turner. An integrative theory of intergroup conflict. In S. Worchel and W. Austin, editors, *The Social Psychology of Intergroup Relations*. Brooks/Cole, Monterey, CA, 1979.

[40] Y. Tang, Y. Zhang, N. Chawla, and S. Krasser. SVMs modeling for highly imbalanced classification. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, 39(1):281–288, 2009.

[41] K. Veropoulos, C. Campbell, and N. Cristianini. Controlling the sensitivity of support vector machines. In *Proceedings of the International Joint Conference on AI*, pages 55–60, 1999.

[42] D. Zhou, O. Bousquet, T. Lal, J. Weston, and B. Schölkopf. Learning with local and global consistency. In *Advances in Neural Information Processing Systems 16: Proceedings of the 2003 Conference*, pages 595–602, 2004.

[43] X. Zhu, Z. Ghahramani, and J. Lafferty. Semi-supervised learning using gaussian fields and harmonic functions. In *Proceedings of ICML*, pages 912–919. ACM, 2003.