Improving Borrowing Behavior Through Social Media Analysis

Rich Colbaugh **Kristin Glass**

Santa Fe, USA and Pyrgadikia, Greece

1. Introduction

Studies of borrowing have tended to assume individuals act rationally and therefore have emphasized improved access to information and increased financial literacy as ways to help borrowers make better decisions. More recent research incorporates “non-rational” drivers of behavior, such as emotions and peer effects, into borrowing models [e.g. 1-5]. These models provide a more complete and accurate view of borrowing behavior and offer valuable insights, but also have significant limitations:

- the models are descriptive, rather than predictive/prescriptive;
- they require data on individual-level outcomes, the collection of which is typically labor-intensive (e.g. using surveys) and gives rise to privacy concerns;
- model estimation relies on standard econometric techniques, which may not be adequate in light of the nonlinear relationships and large-scale, high-dimensional, noisy datasets involved [6].

This paper addresses these limitations by proposing to help individuals anticipate and avoid harmful borrowing results with an approach that combines three novel elements. First, the methodology *employs user-generated social media content* as the primary data source. Observe that social media data is easy to collect and contains volunteered information on user goals, experiences, perceptions, and feelings, which make it an attractive alternative to surveys and similar instruments. Second, *individual-level prediction models are learned from data on aggregate outcomes* via a new machine learning algorithm, thereby eliminating the need to access data on individual loan. Finally, these learned models facilitate *prescriptive analysis,* in which the drivers of poor borrowing behavior are identified and strategies aimed at reducing the negative impact of such borrowing are designed.

The proposed prediction/assistance process consists of the following four steps:

- collect user-generated social media content and analyze this data to infer individuals’ demographic attributes and emotional states;
- combine the social media-derived attributes with aggregate data on borrowing (e.g. at the city level) to learn a model capable of accurately predicting loan outcomes, and use this model to predict the likelihood that a contemplated borrowing decision will yield a negative result;
- leverage the borrowing model to identify the primary drivers of the predicted detrimental borrowing behavior and quantify relative driver importance;
- design an evidence-based behavior-change strategy informed by the inferred behavior-drivers, and deploy the scheme in real-time through a personalized “app” intended to help individuals avoid undesirable loan outcomes.

Figure 1 depicts a high-level schematic of the basic analytic flow.

Figure 1. High-level view of proposed analysis process. First, social media data is analyzed to characterize the demographic attributes and emotional state of a given user. These features serve as inputs to a model which predicts the likelihood of poor borrowing outcomes and identifies the drivers of the undesirable outcomes. Finally, this information is used to give personalized borrowing advice in real-time via a mobile "app".

In the remainder of this paper, we describe the proposed analysis process in more detail, highlight its key innovative elements, and demonstrate its efficacy through case studies involving five types of poor borrowing activity. The presentation is organized as follows. Section 2 summarizes the overall analysis process. Candidate borrowing behavior drivers are identified in Section 3 via computational analysis of online conversations about borrowing. Sections 4 through 6 introduce three central components of the methodology: learning individual-level models from aggregate data, identifying the main drivers of poor borrowing behavior, and reducing the likelihood of negative loan outcomes. In Section 7, the proposed analysis process is applied to five problematic borrowing behaviors: credit card delinquency, borrowing at excessive interest rate, auto loan delinquency, student loan delinquency, and having debt in collections. Finally, Section 8 offers concluding remarks. It should be mentioned that the material reported in this paper derives from two separate but complementary research projects, supported by the Think Forward Initiative (http://www.thinkforwardinitiative.com/) and the US National Science Foundation, and that the results are combined here for clarity of exposition.

2. Overall Analysis Process

The proposed approach to helping individuals anticipate and avoid harmful borrowing exploits social media content and open-source aggregate loan data to generate personalized financial advice, and consists of the sequence of predictive and prescriptive analysis steps shown in Figure 1. To use the system, an individual need only grant access to her public social media data – no financial (or other) data must be entered into the system or otherwise made available. Consequently, the methodology complements other financially-oriented services, is convenient for individuals to adopt and use, and reduces vulnerabilities associated with to private information.

The main steps of the analysis process are now summarized.

Social media analysis. In this step, a given individual's social media posts are collected and analyzed in order to characterize her demographic and emotion traits (e.g. age, ethnicity, political and religious orientation, level of education, feelings of happiness/satisfaction/anger/worry), . The posts are gathered and preprocessed using methods detailed in [7], resulting in a "bag-of-words" model for the content contained in each post. This content is then mapped to demographic/emotion features by applying tools developed in [7-10]. These tools enable a range of personal attributes to be inferred solely from user-generated content. Moreover, unlike standard supervised learning techniques, these tools can be applied *without labeled examples,* by leveraging existing domain lexicons via "lightly-supervised" learning [7-10]. However, to implement these tools it is necessary to specify the attributes to be inferred, that is, the candidate behavior drivers related to poor borrowing (e.g. level of education, conscientiousness). Unfortunately, relatively little is known about these potential drivers. Thus a key phase in this project is the identification of *possible* borrowing behavior drivers through analysis of social media conversations around borrowing; this analysis is described in Section 3. The extent to which these candidate drivers do in fact drive borrowing behavior can then be assessed via predictive analysis (see Section 5).

The inputs to this step are the social media posts generated by the system-user together with a set of candidate borrowing behavior drivers, and the outputs are categorical (e.g. ethnicity) or numerical (five-point Likert scale) values for the candidate attributes as inferred from the social media data (see [7-10] for further details concerning analytic output).

Loan outcome prediction. In the next step of the analysis, the target individual's demographics and emotional state are used to predict the likelihood that the loan presently being considered will have an undesirable outcome (e.g., default). This likelihood is estimated with a borrowing outcome model constructed through machine learning [11]. Crucially, and in sharp contrast to supervised or semi-supervised "early warning" models derived from examples of individual-level outcomes [12-14], this prediction model is learned from *aggregate* outcomes, such as the frequency of credit card delinquency in a set of cities. While the capability to learn individual-level models from aggregate data is clearly desirable, very little has been done to address this challenging problem [15]. Therefore, an important phase of the research was the creation of a novel algorithm for learning individual level models from aggregate (label) data. In particular, the algorithm makes predictions by combining unsupervised feature extraction, aggregate-based modeling, and optimal integration of aggregate-level and individual-level information; the methodology is summarized in Section 4. (Note that an application of this methodology in the human health domain is given in [16].)

The inputs to this step are values for the candidate attributes of the individual of interest, inferred from the social media data in the preceding step, and the output is the probability that the contemplated loan will result in an undesirable outcome.

Drivers of poor borrowing. The third step in the analytic flow is to determine the drivers of harmful borrowing behavior using the prediction model learned in the preceding step. These drivers are identified by assuming that the most predictive features in a good prediction model are the drivers of the behavior being predicted. Specifically, given a borrowing setting of interest, we evaluate the performance of the loan outcome prediction model corresponding to this setting and, if performance is satisfactory, we conclude that the model's predictive features are behavior-drivers. Note that this approach is consistent with data-driven causal inference [17] and – though causality is not proved – is found to work well in practice. Predictive features are identified and ranked by computing and integrating a set of "feature importance" metrics for the learned model; the method is summarized in Section 5.

The input to this step is the loan outcome prediction model learned in the preceding step, and the output is a list of drivers of poor borrowing, ranked by their relative significance.

Reducing negative outcomes. Finally, the results obtained in the preceding analysis steps are integrated to support the design and deployment of a system aimed at helping individuals avoid the predicted undesirable outcomes. Our current emphasis is on evidence-based behavior-change methods, such as stimulus control [18], delivered through a mobile “app”. This strategy enables scalable execution of user-interaction that is personalized, timely, data-driven, and convenient, qualities which have been shown to increase the probability of success. Additionally, the concept enables non-rational drivers of borrowing behavior to be combined with financial data (when available) to formulate guidance during a purchase journey, a capability which is currently not available to consumers.

The present concept is briefly summarized in Section 6. It should be noted, though, that as our initial solution has been selected by the Think Forward Initiative to be developed into a prototype solution during an accelerator program at TFI Labs, it likely to change and progress in the coming months.

3. Analyzing Social Media Conversations

There is considerable interest in analyzing social media data to infer user attributes, and as a result there now exist several analytic tools with which to predict demographic, emotion, and personality attributes from user-generated content [7-10,19]. For example, we have derived methods for accurately estimating emotional state (e.g. happiness, anger, worry) and personal characteristics (e.g. level of education, religiosity, political orientation, income, ethnicity) from Twitter posts [7-9,19]. As indicated in Section 2, the first step in the proposed analysis process is to analyze the social media activity of a given user and infer attributes and emotional states which may be predictive of borrowing outcomes (see Figure 1). To apply the tools described in [7-10,19] to this task, it is necessary to specify the attributes to be inferred, that is, the candidate behavior drivers associated with poor borrowing, and unfortunately little is known about these drivers. Therefore a key phase in this research project is identifying possible borrowing drivers through analysis of social media conversations around borrowing; the approach adopted to accomplish this identification is now summarized.

To conduct the analysis, we applied unsupervised natural language processing techniques to uncover important online conversation themes around borrowing. Data was collected through web-crawls targeting personal borrowing discussions. These crawls were launched from a heterogeneous set of initiating seed sites (e.g. spanning diverse audience demographics), implemented focusing content filters of the form ‘personal finance’ AND [‘borrowing’ OR ‘debt’ OR …] [7], and returned approximately 10K pertinent conversation threads. Each thread was modeled as a bag-of-words document, with the set of all threads forming the corpus of documents D to be analyzed. The main topics underlying the conversation threads in D were then identified through Latent Dirichlet Allocation (LDA), a generative probabilistic modeling procedure devised for discrete data such as natural language [20]. More precisely, LDA learns the “latent” topics (which are probability distributions over words) and the mixture of topics for each document, by maximizing the likelihood of the corpus D [20].

Applying LDA to the corpus of conversation threads D resulted in the discovery of the following set of topics, ordered by significance:

- negative emotions about debt;
- negative emotions around overspending;

- credit card payments/delinquency/complaints/questions/advice;
- questions and advice concerning borrowing/payment/delinquency (general);
- paying for college/student loans (including relationships and emotional factors);
- automobiles and paying for them (including relationships and emotional factors);
- vacations and paying for them (including relationships and emotional factors);
- budgeting tools.

Example posts taken from threads discussing the first two topics, negative emotions about debt and negative emotions around overspending, are displayed in Figure 2.

It is stressed that the topics detected via this analysis are used to generate *candidate* drivers of harmful borrowing, and that the validity of the candidates is evaluated through predictive analysis with particular loan types (see Sections 5 and 7).

I became increasingly unhappy and resentful about my husband‘s unemployment, and I just continued spending through all of it. And a lot of what I spent money on was unnecessary .... within about 6 months I had over $10,000 in credit card debt again.

I just got a buzz out of it all the time, but then before I knew it I was $20k in debt on my four credits cards.

Boredom has a lot to do with it. Or if I'm feeling fed up, you know, I have sometimes just decided I'll go and spend some money because I'm fed up.

Because I know that whatever, if I'm really feeling pissed-off, I can go shopping and that will cheer me up.

It never feels like I'm actually spending money if I use credit cards. I just hand the card over and I just think, well, I'll worry about it when the bill comes.

What happened to me with my marriage is I was being controlled and I‘d be mad at him and I‘d go out shopping. I would feel like if I was in the store shopping with my own money to spend I was just like all the other people in the store that had money to spend.

For me I was just angry inside and I didn‘t know the reason. So I was filling myself with shopping and credit cards.

Figure 2. Sample social media posts taken from the topics ‘negative emotions about debt’ and ‘negative emotions around overspending’.

4. Learning from Aggregate Data

Constructing models to predict individual-level (IL) behavior via standard machine learning or econometric techniques requires that large numbers of IL training examples be labeled in accordance with the outcome to be predicted [11]. In financial applications, labeling the outcomes of individuals (e.g. whether an auto loan taken by a person is repaid on time) is labor-intensive, often demands domain expertise, and gives rise to privacy concerns. Consequently, the need for IL labeling represents a significant obstacle to deriving empirically-grounded models for financial behavior. To overcome this obstacle, we have developed a new machine learning methodology which enables accurate IL prediction models to be learned from *aggregate* labels, corresponding to group rather than individual outcomes (e.g. credit card delinquency rates for each state in the US rather than the delinquency status of individual card-holders). This section of the paper offers a summary of the new learning process.

4.1 Problem formulation

We begin by defining the problem of interest. Given *aggregate* information about the (hidden) labels of individuals (e.g. the fraction of people who have defaulted on their student loans in each of the US states), the goal is to learn a model which accurately predicts *individual-level* labels (whether a particular person will repay their student loan). For simplicity it is assumed that the prediction task is binary classification (e.g., loan default or not), but the approach can be extended to multi-class classification and regression problems.

To state the problem quantitatively we introduce two data models, corresponding to the two levels of analysis:

- instance-level: $D_I = \{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)\}$, where $\mathbf{x} = [x_1, \ldots, x_d]^T$ is the feature vector and (crucially) the labels $y_i \in \{-1,+1\}$ are unknown (even with the training data);
- aggregate-level: instances occur in bags $\{\mathbf{x}_i \mid i \in B_j\}$, $j \in \{1, \ldots, m\}$; each bag is modeled by the mean $\mathbf{z}_j$ of the instance feature vectors in that bag and is labeled with the bag's fraction of positive instances f_j (assumed known), yielding the labeled aggregate data $D_A = \{(\mathbf{z}_1, f_1), \ldots, (\mathbf{z}_m, f_m)\}$.

The problem is then: given D_A, learn a model which predicts the probability p_i that any instance $\mathbf{x}_i \in D_I$ has label $y_i = +1$.

4.2 Learning method

The proposed approach to learning IL prediction models from aggregate data consists of three steps: feature extraction, aggregate-based prediction, and clustering-informed prediction refinement (see Figure 3). We now describe each step in the process and then summarize the complete model by specifying the ensemble learning from aggregate labels (ELAL) algorithm.

Feature extraction. The first step in the prediction process is to transform the given instance representation $\mathbf{x}^o \in \Re^{do}$ to the form $\mathbf{x} = [x_1, \ldots, x_d]^T$ used in data model D_I, where feature vector $\mathbf{x}$ better captures the underlying structure of the data [11]. We employ two techniques to learn the transformation $\mathbf{x}^o \rightarrow \mathbf{x}$: Latent Dirichlet Allocation (LDA) [20] and deep learning via stacked autoencoders (SAE) [21]. In each case, it is assumed that the data distribution concentrates near a low-dimensional manifold, and the objective is to identify "latent" variables that model this manifold.

For example, analysis of social media often focuses on the content of posts, and it is common to model each post (or collection of user-posts) as a bag of words $\mathbf{x}^o \in \Re^{|V|}$, where the entries of $\mathbf{x}^o$ are the frequencies with which the words in vocabulary V appear in the post(s). The latent features identified by LDA then have a natural interpretation as the topics being discussed in the posts (see Section 3).

Alternatively, with SAE the goal is to learn a parsimonious encoding of $\mathbf{x}^o$, $\mathbf{x} = f_\theta(\mathbf{x}^o)$, which permits accurate decoding $\mathbf{x}_d = g_\theta(\mathbf{x}) \approx \mathbf{x}^o$, where f_θ and g_θ are encoding and decoding functions, respectively, and θ is the vector of parameters to be learned [21]. Note that both LDA and SAE learn feature representations in an unsupervised fashion, so each can be implemented directly with data D_I (recall that instance-level labels y_i are unknown, so unsupervised learning is necessary).

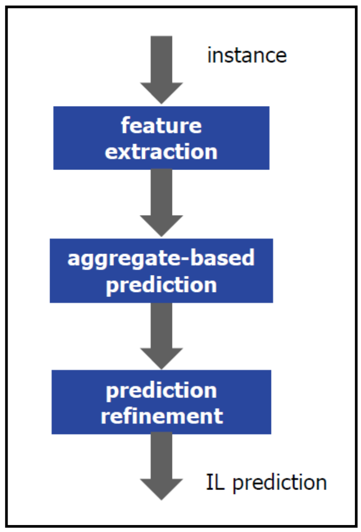

Figure 3. Illustration of the proposed approach to making IL predictions using models learned from aggregate data.

Aggregate-based prediction. The second step in the prediction process is to leverage the aggregate data D_A to induce a model which enables preliminary IL predictions to be made. As the aggregate data D_A is labeled, this prediction model can be obtained through a two-stage (lightly) supervised learning procedure:

- aggregate-level regression: given labeled aggregate data $D_A = \{(\mathbf{z}_1, f_1), \ldots, (\mathbf{z}_m, f_m)\}$, learn an ensemble of decision trees [11] regression model $\mathbf{f}_r$: Z $\rightarrow$ F that accurately predicts the fraction of positive instances f_* in a new (unseen) bag B_*;
- instance-level prediction: apply the aggregate-level model $\mathbf{f}_r$ to the instances $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ in D_I to form predictions $\mathbf{p}_0 = [p_{01}, \ldots, p_{0n}]^T$, where p_{0i} is the predicted probability that instance $\mathbf{x}_i$ has label $y_i = +1$.

Because the predictions $\mathbf{p}_0$ are derived using a model learned on the mean behaviors $(\mathbf{z}_j, f_j)$ of bags of instances, they are on average good estimates for the probabilities that the instances $\mathbf{x}_i$ have positive labels. However, prediction quality may have large variance (e.g. if the number of bags is small), motivating our interest in a prediction-refinement step.

Prediction refinement. The third step in the prediction process involves refining preliminary predictions $\mathbf{p}_0$ to final predictions $\mathbf{p} = [p_1, \ldots, p_n]^T$ (p_i is the predicted probability that $\mathbf{x}_i$ has label $y_i = +1$). The idea behind the refinement process is that if instances $\mathbf{x}_k$, $\mathbf{x}_l$ are contained in the same bag and are "similar", then they should have similar labels. This reasoning suggests that IL data, if informative concerning instance similarity, may be helpful in improving predictions $\mathbf{p}_0$.

Let L_j=I–C_j, where I is the identity matrix and C_j is a similarity matrix computed via ensemble clustering [22] on bag B_j. Specifically, we construct matrix C_j so that its (k,l) entry is equal to the number of times B_j's instances $\mathbf{x}_k$, $\mathbf{x}_l$ are assigned to the same cluster by members of the ensemble. Final predictions $\mathbf{p}$ are formed by optimally balancing the goals of maintaining agreement with initial predictions $\mathbf{p}_0$ and achieving within-cluster label similarity:

$$\min_{\mathbf{p}} \{ \lambda \, \| \mathbf{p} - \mathbf{p}_0 \|^2 + (1-\lambda) \, \Sigma_{j \in \{1,\ldots,m\}} \, \mathbf{p}_j^T \, L_j \, \mathbf{p}_j \qquad (1)$$

subject to the constraint that $p_i \in [0,1] \; \forall i$, where $\mathbf{p}_j$ is the subset of predictions $\mathbf{p}$ that correspond to bag B_j and hyperparameter $\lambda \in [0,1]$ reflects the relative expected predictive value of aggregate-level label data and instance-level clustering.

It is seen that the predictions $\mathbf{p}$ generated by minimizing (1) incorporate information from three sources: unsupervised IL feature learning, supervised aggregate-level learning (via $\mathbf{p}_0$), and unsupervised IL clustering (via the L_j). The optimization (1) can be accomplished, independently for each bag, by iterating the following formula over index i until convergence (which is guaranteed [22]):

$$\mathbf{p}^{i+1} = \lambda \, \mathbf{p}_0 + (1-\lambda) \, C_{norm} \, \mathbf{p}^i, \quad \mathbf{p}^0 = \mathbf{p}_0. \qquad (2)$$

In (2), C_{norm} is the normalized version of C obtained by transforming C into a symmetric probability matrix [22]. This solution is efficient to compute, allowing large-scale problems to be investigated.

We summarize this discussion by sketching the proposed algorithm for making IL predictions.

Algorithm ELAL:

1. Learn feature representation $\mathbf{x}$ for instance $\mathbf{x}^o$ which captures the underlying structure of the data (e.g. using LDA [20] or SAE [21]).
2. Use aggregate-level labeled data D_A to learn an ensemble of decision trees regression model $\mathbf{f}_r$ that accurately predicts the fraction of positive instances f_* in a new (unseen) bag B_*.
3. Compute preliminary predictions $\mathbf{p}_0$ for all the instances in D_I using $\mathbf{f}_r$.
4. Perform prediction refinement $\mathbf{p}_0 \rightarrow \mathbf{p}$ by optimizing (1) using the iteration (2).

The performance of Algorithm ELAL is now illustrated through a test case: income prediction.

4.3 Test: Income prediction

Consider the task of accurately identify low-income individuals based only on the content of their Twitter posts. The data employed to learn the prediction model and evaluate its performance consists of the Twitter posts and income levels of 5191 individuals, 500 with low income and 4691 with "normal" income, collected in 2014 [23]. These individuals are grouped into 20 geo-location-based bags using the scheme described in [24]; this aggregation procedure simulates the common situation where financial or economic statistics are publicly-available for geographic regions (e.g. proportion of low-income people living in each US county) but not for individuals.

Algorithm ELAL is implemented to solve this prediction problem in the following way:

- feature extraction: LDA learns 200 topics being discussed in the Twitter posts (where the number of topics was selected through perplexity analysis [20]), and the extent to which user i posts about each of these topics defines the d = 200 elements of feature vector $\mathbf{x}_i$;
- aggregate-based prediction: an ensemble of 1000 regression trees is learned on labeled aggregate data D_A where, for each bag B_j, $\mathbf{z}_j$ is the mean topic-based Twitter feature vector and f_j is the fraction of low-income individuals in that bag; the learned regression model $\mathbf{f}_r$ is then applied to instances $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ in D_I to form preliminary IL predictions $\mathbf{p}_0$;
- prediction refinement: the refinement $\mathbf{p}_0 \rightarrow \mathbf{p}$ is accomplished by optimizing (1) using (2) with λ = 0.3 (set by rough tuning on a small held-out validation set).

For comparison purposes, two other prediction models are tested: *naïve*, which simply predicts the bag's majority class for all individuals in a given bag, and *instance-level*, which is a state-of-the-art random forest (RF) classifier [11] trained on 3000 labeled instances (these instances are modeled with the same topic-based features used by Algorithm ELAL). Predictive performance is measured with class-average (CA) accuracy, estimated through cross-validation. (Recall CA accuracy is the average of sensitivity and specificity, and is an informative measure of performance when there exists significant class-imbalance [11].)

The results obtained in this study are displayed in Figure 4. It can be seen that Algorithm ELAL, which has access to *no* IL labels, achieves CA accuracy (90.1%) comparable to a state-of-the-art classifier trained with 3000 labeled instances (90.6%). (We choose 3000 labeled instances to train the RF classifier to achieve results similar to those attained with Algorithm ELAL; using fewer labeled examples yields lower accuracy.)

5. Identifying Drivers of Borrowing Behavior

The third step in the proposed approach to helping individuals anticipate and avoid undesirable borrowing outcomes is to determine the drivers of harmful borrowing behavior (see Section 2 and Figure 1 for context). These drivers are identified by assuming that the most predictive features in a good prediction model are the drivers of the behavior being predicted. Specifically, given a borrowing setting of interest, we evaluate the performance of the loan outcome prediction model built for this setting and, if performance is satisfactory, we conclude that the model's predictive features are behavior-drivers. This approach is consistent with data-driven causal inference [17] and – though causality is not proved – it is found to work well in practice.

Predictive features are identified and ranked by computing and integrating a set of "feature importance" metrics for the learned model. Evaluation of feature predictive power is accomplished using the procedure detailed below.

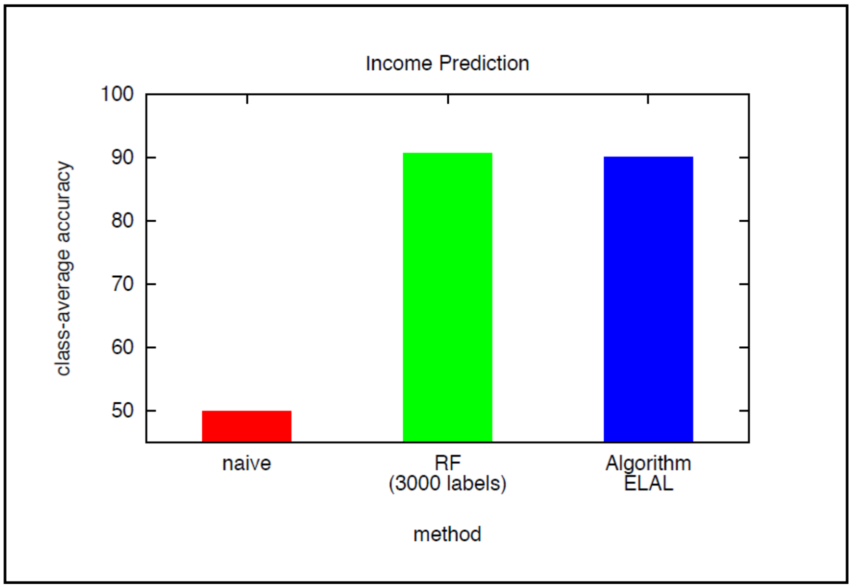

Figure 4. Results for income prediction test. The plot compares class-average accuracies obtained with three prediction methods: naïve (red), individual-level learning trained on 3000 labeled instances (green), and Algorithm ELAL (blue).

Given a set of user attributes inferred from social media data (see Step 1 in Figure 1) and aggregate data for borrowing outcomes corresponding to the loan type of interest:

- Learn a good prediction model for the target domain using Algorithm ELAL.
- Estimate the predictive power of all feature by employing each of four measures:
    - Breiman feature importance (each feature's predictive power is estimated based on the impact of randomly permuting the feature's values over a test set) [25];
    - leave-one-out prediction (each feature's predictive power is estimated based on the performance of the prediction model built using all but that feature) [11];
    - use-only-one prediction (each feature's predictive power is estimated based on the performance of the prediction model built using only that feature) [11];
    - single-feature predictability (each feature's predictive power is estimated based on the magnitude of transfer entropy between that feature and the label) [26].
- Combine the four measures of predictive power via Borda ranking [27] to obtain the final rank-ordered list of predictive features.

The output of this procedure is a list of features found to have predictive power, ordered according to power. The method is illustrated in Section 7, where behavior drivers are determined for five diverse loan types.

6. Reducing Negative Outcomes

Once a good prediction model is learned and the model detects that a contemplated borrowing transaction is likely to produce a poor outcome, and the drivers associated with the poor result are identified, attention can be turned to helping the individual avoid the undesirable outcome (see Section 2 and Figure 1 for context). We are presently working on the design and deployment of a system which can provide this assistance. Our current focus is on evidence-based behavior-change methods, such as stimulus control [18], delivered through a mobile "app". This strategy enables scalable implementation of user-interaction that is personalized, timely, data-driven, and convenient, qualities which have been shown to increase the probability of success.

Our initial concept was "a recommender service for helping you decide if you can afford a product', and has been selected by the Think Forward Initiative to be developed into a prototype solution during an accelerator program at TFI Labs. This prototype is likely to progress rapidly in the coming months, so here we simply list a few characteristics of the initial concept which differentiate it from existing solutions.

The "app"

- enables non-rational drivers of borrowing behavior (e.g. emotional state) to be combined with an individual's aggregate financial data (e.g. bank account balance, credit card balances) to deliver personalized-yet-neutral, relevant, and convenient-to-access information throughout a purchase journey;
- generates simple, meaningful recommendations upon which the user can base a purchase decision anf that works for them both financially and emotionally;
- presents options to purchase, step away, or work towards making the purchase at a later stage (see Figure 5);
- will be available through a variety of channels (e.g. mobile device, Web browser) to facilitate its use throughout the purchase journey; for example, it is expected that the information supplied by this decision-support system may be useful for activities ranging from early research in preparation for a large purchase all the way to smaller "impulse buys".

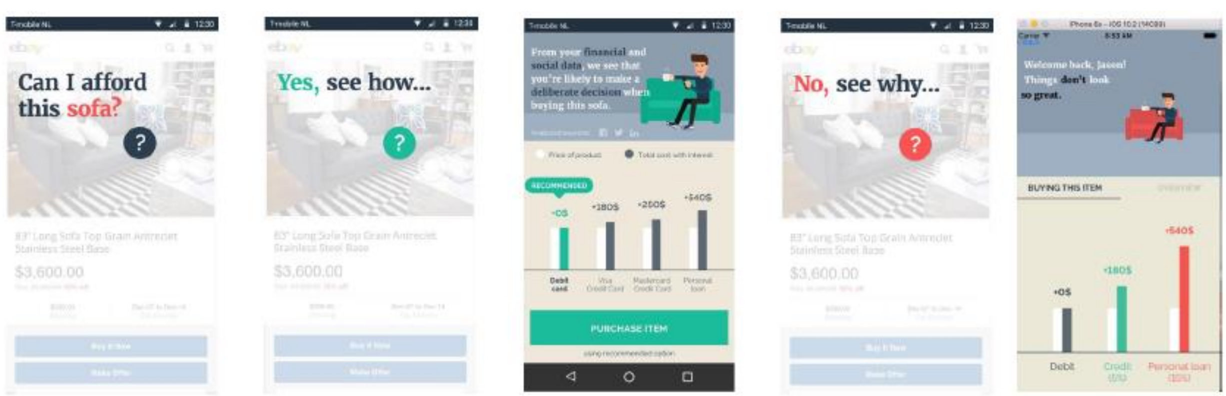

Figure 5. Illustrative screen shots from an initial prototype of the "app".

7. Case Studies

We now present several case studies which illustrate the performance of the proposed analytics methodology. These studies address the tasks of providing early warning of harmful borrowing outcomes and identifying the behavior-drivers underlying these outcomes. The case studies investigate five diverse loan types and outcomes:

- credit card delinquency;
- debt in collections;
- auto loan delinquency;
- student loan delinquency;
- high interest borrowing.

In each case study, the source of IL social media data is a large set of Twitter posts collected in 2013 [28]. Specifically, the dataset consists of 5.7M Tweets, geo-located at the US state-level, with geo-location performed according to the scheme outlined in [24]. All borrowing outcome data is aggregated, also at the state-level; the sources of aggregate data are cited in the respective case studies (data is not held by the authors). The task is then to learn a model from the IL Twitter data and aggregate loan outcomes which enables accurate prediction of *individual* loan outcomes (see Section 4).

Evaluating the performance of the resulting learned models is complicated by the fact that IL outcomes are not available for the case studies. (For example, attempts to gain access to such data in collaboration with the Think Forward Initiative, as a sponsor of the research, were unsuccessful.) Thus we adopt a two-component evaluation process:

- aggregate-level performance: the aggregate-level prediction model learned in Step 2 of Algorithm ELAL, which predicts state-level borrowing outcomes from aggregate Twitter posts, is evaluated out-of-sample via cross-validation, with CA accuracy as the metric;
- individual-level performance: the IL prediction model learned using Algorithm ELAL is evaluated by estimating the area under the ROC curve (AUC) [11] for the model using the procedure given in [15]; crucially, this evaluation technique requires only aggregate-level labels to estimate IL AUC and so can be implemented in the present setting.

Note that, because standard predictive modeling methods are not capable of learning IL models from aggregate data, no comparison can be made to “baseline” techniques in the case studies. However, we do compare prediction performance obtained using different classes of features (e.g. financially-oriented, social media-derived).

7.1 Credit card delinquency

Task: Learn a model which predicts 2014 IL credit card (CC) delinquency using 2013 Twitter data, with learning based upon aggregate state-level delinquency rates, and use the model to identify the drivers of IL delinquency behavior.

Datasets: State-level CC delinquency rates [29] and IL geo-located Twitter data [28].

Aggregate-level performance: Aggregate-level prediction results are presented in Figures 6 and 7. Figure 6 displays illustrative results generated in the first step of the analytic process, in which emotions and demographics are inferred from Twitter posts. Figure 7 shows the CA accuracy of the aggregate-level prediction model learned in Step 2 of Algorithm ELAL for a variety of feature sets.

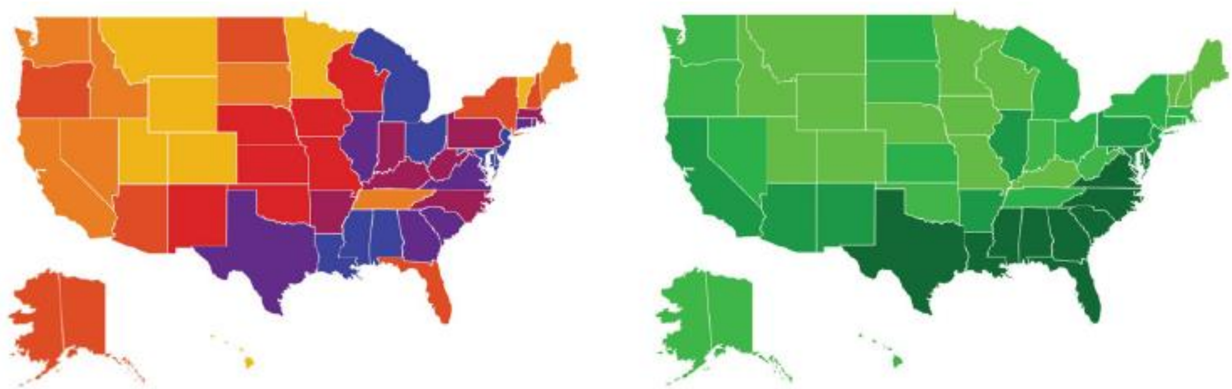

Figure 6. Illustrative "heat maps" of emotions inferred from Twitter posts, aggregated to US state-level for convenience of display. Left map is estimated happiness level (warmer colors are more happy) and right map is estimated anger level (darker colors are more angry).

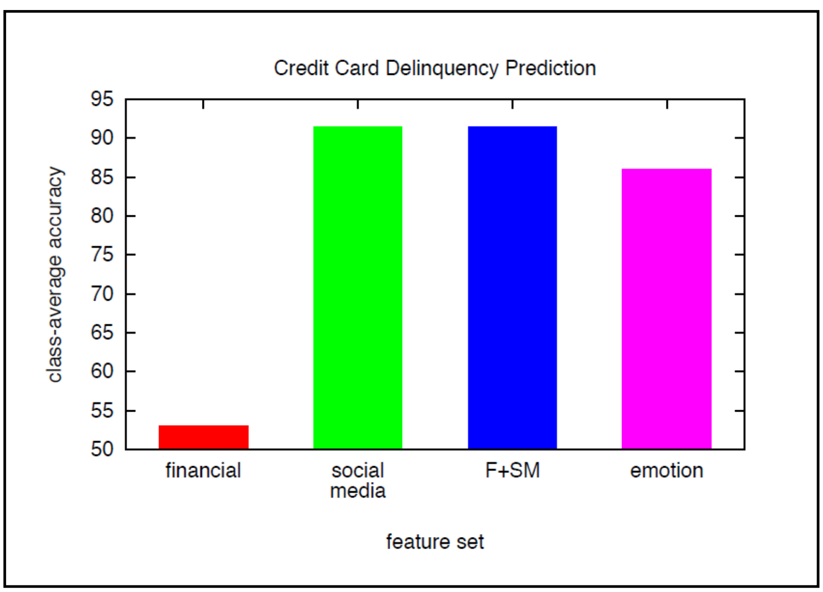

Figure 7. Performance of aggregate-level prediction model. All results use models learned via Step 2 of Algorithm ELAL, but each employs a different feature set: income data alone (red), Twitter posts alone (green), income plus Twitter (blue), and inferred emotions alone (magenta).

Individual-level performance. IL prediction results are as follow. Estimated AUC for Algorithm ELAL learned on Twitter data alone (e.g. no financial data) is 96.7%. The main drivers of CC delinquency are, in rank order of importance:

- emotion, especially worry, anger, and unhappiness;
- level of education;
- religiosity;
- cultural factors.

7.2 Debt in collections

Task: Learn a model which predicts 2014 IL debt in collections using 2013 Twitter data, with learning based upon aggregate state-level rates of debt in collections, and use the model to identify the drivers of IL debt behavior.

Datasets: State-level rates of debt in collections [29,30] and IL geo-located Twitter data [28].

Aggregate-level performance: Aggregate-level prediction results are presented in Figure 8. This figure shows the CA accuracy of the aggregate-level prediction model learned in Step 2 of Algorithm ELAL for a variety of feature sets.

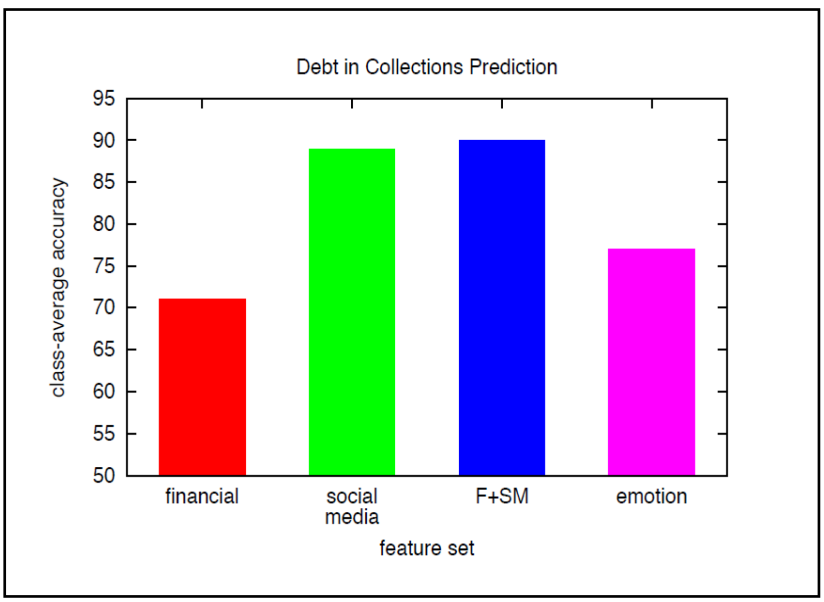

Figure 8. Performance of aggregate-level prediction model. All results use models learned via Step 2 of Algorithm ELAL, but each employs a different feature set: income data alone (red), Twitter posts alone (green), income plus Twitter (blue), and inferred emotions alone (magenta).

Individual-level performance. IL prediction results are as follow. Estimated AUC for Algorithm ELAL learned on Twitter data alone (e.g. no financial data) is 90.3%. The main drivers of debt in collections are, in rank order of importance:

- level of education;
- emotion, especially life satisfaction;
- political orientation;
- income.

7.3 Auto loan delinquency

Task: Learn a model which predicts 2015 IL auto loan delinquency using 2013 Twitter data, with learning based upon aggregate state-level delinquency rates, and use the model to identify the drivers of IL auto loan delinquency behavior.

Datasets: State-level auto loan delinquency rates [29,31] and IL geo-located Twitter data [28].

Aggregate-level performance: Aggregate-level prediction results are presented in Figure 9. This figure shows the CA accuracy of the aggregate-level prediction model learned in Step 2 of Algorithm ELAL for a variety of feature sets.

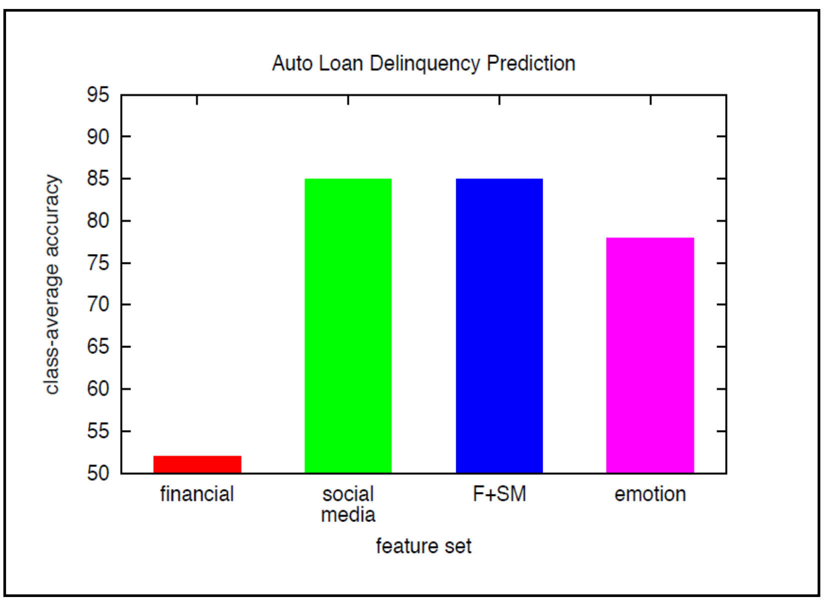

Figure 9. Performance of aggregate-level prediction model. All results use models learned via Step 2 of Algorithm ELAL, but each employs a different feature set: income data alone (red), Twitter posts alone (green), income plus Twitter (blue), and inferred emotions alone (magenta).

Individual-level performance. IL prediction results are as follow. Estimated AUC for Algorithm ELAL learned on Twitter data alone (e.g. no financial data) is 88.5%. The main drivers of auto loan delinquency are, in rank order of importance:

- emotion, especially worry and anger;
- cultural factors;
- level of education;
- religiosity.

7.4 Student loan delinquency

Task: Learn a model which predicts 2014 IL student loan delinquency using 2013 Twitter data, with learning based upon aggregate state-level delinquency rates, and use the model to identify the drivers of IL student loan delinquency behavior.

Datasets: State-level student loan delinquency rates [29,32] and IL geo-located Twitter data [28].

Aggregate-level performance: Aggregate-level prediction results are presented in Figure 10. This figure shows the CA accuracy of the aggregate-level prediction model learned in Step 2 of Algorithm ELAL for a variety of feature sets.

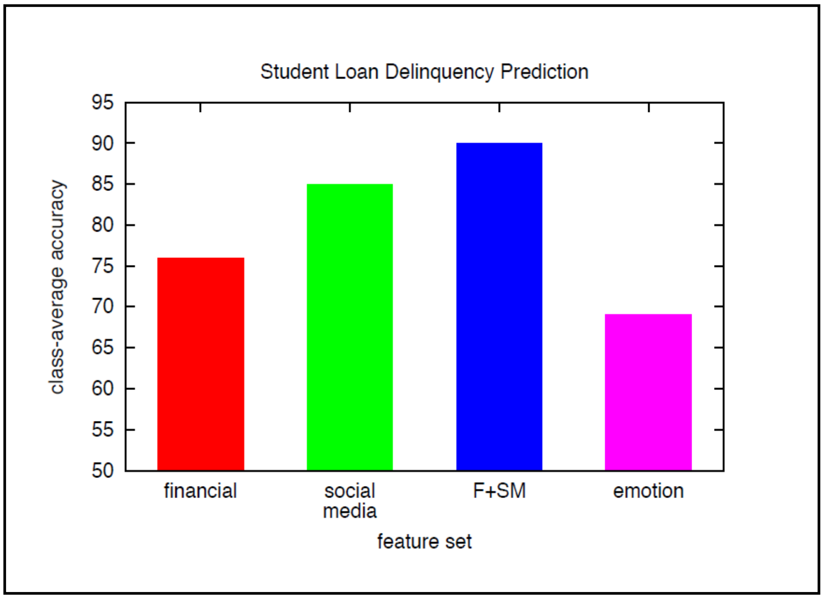

Figure 10. Performance of aggregate-level prediction model. All results use models learned via Step 2 of Algorithm ELAL, but each employs a different feature set: income data alone (red), Twitter posts alone (green), income plus Twitter (blue), and inferred emotions alone (magenta).

Individual-level performance. IL prediction results are as follow. Estimated AUC for Algorithm ELAL learned on Twitter data alone (e.g. no financial data) is 90.2%. The main drivers of student loan delinquency are, in rank order of importance:

- level of education;
- income;
- political orientation;
- emotion, especially happiness.

7.5 High interest borrowing

Task: Learn a model which predicts 2014 IL high interest borrowing events using 2013 Twitter data, with learning based upon aggregate state-level rates of peer-to-peer (P2P) borrowing (as a proxy), and use the model to identify the drivers of high interest borrowing behavior.

Datasets: State-level per capita P2P borrowing rates [29,33] and IL geo-located Twitter data [28].

Aggregate-level performance: Aggregate-level prediction results are presented in Figure 11. This figure shows the CA accuracy of the aggregate-level prediction model learned in Step 2 of Algorithm ELAL for a variety of feature sets.

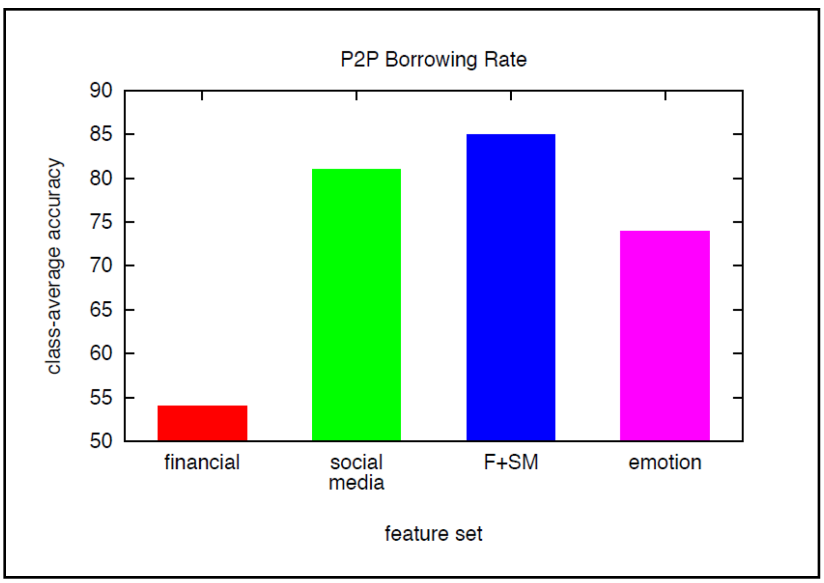

Figure 11. Performance of aggregate-level prediction model. All results use models learned via Step 2 of Algorithm ELAL, but each employs a different feature set: income data alone (red), Twitter posts alone (green), income plus Twitter (blue), and inferred emotions alone (magenta).

Individual-level performance. IL prediction results are as follow. Estimated AUC for Algorithm ELAL learned on Twitter data alone (e.g. no financial data) is 89.2%. The main drivers of debt in collections are, in rank order of importance:

- cultural factors;
- emotion;
- religiosity;
- political orientation.

For convenience, the IL predictive modeling results for the case studies are summarized in Figure 12.

Prediction Performance for Various Loan Types		
Loan Type	**AUC**	**Key Predictors**
credit card delinquency	96.7%	emotion, level-of-education
student loan delinquency	90.2%	level-of-education, income
debt in collections	90.3%	level-of-education, emotion
auto loan delinquency	88.5%	emotion, cultural factors
high-interest borrowing	89.2%	cultural factors, religiosity

Figure 12. Summary of the case study results: individual-level prediction performance (AUC) and key predictors/behavior-drivers for each of the five loan types.

8. Discussion

This paper proposes an approach to helping individuals anticipate and avoid harmful borrowing events by exploiting social media content and open-source aggregate loan data to generate personalized financial advice. The analysis underlying the system consists of a sequence of three main predictive/prescriptive steps. First, social media data is analyzed to characterize the demographic attributes and emotional state of a specific user. These features serve as inputs to a model which predicts the likelihood of poor borrowing outcomes and identifies the drivers of the undesirable outcomes. Finally, this information is used to give personalized borrowing advice in real-time via a mobile "app". Importantly, both rational and non-rational influences are incorporated into the analysis, and no data must be entered into the system by the user. Consequently, the methodology complements other financially-oriented services, is convenient for individuals to adopt and use, and reduces security vulnerabilities associated with private information. The efficacy of the approach is demonstrated through case studies involving five diverse types of borrowing activity.

Our research offers concrete ways to improve borrowing behavior. For instance, by providing early warning that a contemplated loan is likely to produce an undesirable outcome, the proposed system can help users anticipate and avoid these outcomes. Identifying the drivers of poor borrowing behavior facilitates design and implementation of behavior-change strategies aimed at decreasing the likelihood of poor borrowing in the future (e.g., via stimulus control [18]); the mobile "app" summarized in Section 6 is one way to deliver such strategies. At a more fundamental level, our finding that emotions in the weeks leading up to a borrowing event are more important than rational considerations gives a new, empirically-grounded perspective on personal borrowing.

Although the present research focuses on borrowing, the methodology and perhaps even some of the specific results are likely to be applicable to additional aspects of personal finance. For example, several other kinds of financial behavior are known or thought to be influenced by emotions, peer effects, and cognitive biases, including savings, investing, budgeting, and insurance. Thus these behaviors seem to be good candidates for analysis using the approach employed in this paper. The methodological innovations presented here may also be of value in other areas of financial and economics research. Leveraging social media as an inexpensive source of volunteered, economics-relevant data may have much wider utility. Furthermore, our new capability to learn individual-level models for financial activity from aggregate data appears to have substantial potential in this sphere.

9. Acknowledgements

Research supported by the Think Forward Initiative and the US National Science Foundation.

10. References

[1] White, B., "Take this house and shove it: The emotional drivers of strategic default", *Arizona Legal Studies,* Discussion Paper 10-17, May 2010.

[2] Willis, L., "The financial education fallacy", *The American Economic Review,* Vol. 101, 2011.

[3] Wang, L. et al., "Demographics, attitude, personality, and credit card features correlate with credit card debt", *Journal of Economic Psychology*, Vol. 32, 2011.

[4] Bjorkegren, D. and D. Grissen, "Behavior revealed in mobile phone usage predicts loan repayment", Preprint, Brown University, July 2015.

[5] Goodstein, R. et al., "Contagion effects in strategic mortgage defaults", *Journal of Financial Intermediation*, Vol. 30, 2017.

[6] Varian, H. "Big data: New tricks for econometrics", *Journal of Economic Perspectives,* Vol. 28, 2014.

[7] Glass, K. and R. Colbaugh, "Web analytics for security informatics", *Proc. IEEE EISI,* Athens, Greece, September 2011.

[8] Glass, K. and R. Colbaugh, "Agile detection of framing rhetoric in social media", *Proc. IEEE ISI,* Washington, DC, USA, June 2012.

[9] Colbaugh, R. and K. Glass, "Analyzing social media content for security informatics", *Proc. IEEE EISI,* Uppsala, Sweden, August 2013.

[10] Mitchell, L. et al., "The geography of happiness: Connecting Twitter sentiment and expression, demographics, and objective characteristics of place", *PLoS ONE,* Vol. 8, 2013.

[11] Hastie, T., R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, Second Edition, Springer, New York, 2009.

[12] Colbaugh, R. and K. Glass, “Early warning analysis for social diffusion events”, *Security Informatics,* Vol. 1, 2012.

[13] Colbaugh, R. and K. Glass, “Semi-supervised ensemble learning for accurate prediction with limited labeled data”, Presentation to the US National Science Foundation, Las Vegas, NM, USA, April 2016.

[14] Colbaugh, R. and K. Glass, “Asymmetry in coevolving adversarial systems”, *Proc. IEEE QRS,* Vienna, Austria, August 2016.

[15] Yu, F. et al., “On learning from label proportions”, arXiv preprint 1402.5902v2, February 2015.

[16] Colbaugh, R. and K. Glass, “Learning about individuals’ health from aggregate data”, *Proc. IEEE EMBC*, Jeju Island, Korea, July 2017.

[17] Pearl, J., “Causal inference in statistics: An overview”, *Statistical Surveys,* Vol. 3, 2009.

[18] Greaves, C. et al., “Systematic review of reviews of intervention components associated with increased effectiveness in dietary and physical activity interventions”, *BMC Public Health,* Vol. 11, 2011.

[19] Colbaugh, R. and K. Glass, “Detecting emerging topics and trends via predictive analysis of ‘meme’ dynamics”, *Proc. IEEE ISI,* Beijing, China, July 2011.

[20] Blei, D. et al., “Latent Dirichlet allocation”, *Journal of Machine Learning Research,* Vol. 3, 2003.

[21] Bengio, Y. et al., “Representation learning: A review and new perspectives”, *IEEE Trans. PAMI,* Vol. 35, 2013.

[22] Colbaugh, R., “Practical predictability of human behavior”, Computer Science Seminar, University College London, UK, June 2015.

[23] Preotiuc-Pietro, D. et al., “Studying user income through language, behavior, and affect in social media”, *PLoS ONE,* Vol. 10, 2015.

[24] Priedhorsky, R. et al., “Inferring the origin locations of Tweets with quantitative confidence”, *Proc. ACM CSCW*, Baltimore, MD, USA, February 2014.

[25] Breiman, L., “Random forests”, *Machine Learning*, Vol. 45, 2001.

[26] Colbaugh, R. et al., “Dynamic information-theoretic measures for security informatics”, *Proc. IEEE ISI,* Seattle, WA, USA, June 2013.

[27] Dwork, C. et al., “Rank aggregation methods for the Web”, *Proc. WWW10,* Hong Kong, May 2001.

[28] Colbaugh, R. and K. Glass, “2013 US Twitter dataset”. Data contains personally-identifiable information are is therefore considered confidential and handled as such.

[29] Federal Reserve Bank of New York, https://www.newyorkfed.org/, accessed 2016.

[30] TransUnion, https://www.transunion.com/, accessed 2016.

[31] Experian, https://www.experian.com/, accessed 2016.

[32] Federal Reserve Bank of St. Louis, https://www.stlouisfed.org/, accessed 2016.

[33] Orchard Platform, https://www.orchardplatform.com/, accessed 2016.