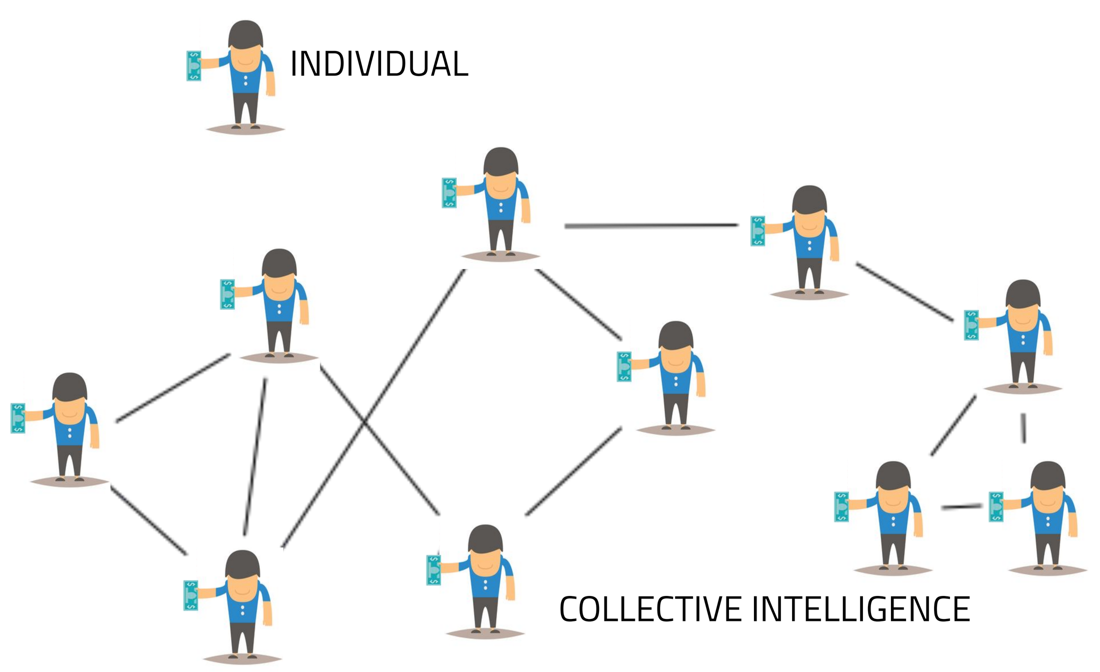
Harnessing Collective Intelligence in Peer-to-Peer (P2P) Lending

Henry K. Dambanemuya and Emőke-Ágnes Horvát*
Northwestern University, Evanston, Illinois
hdambane@u.northwestern and a-horvat@northwestern.edu

INTRODUCTION

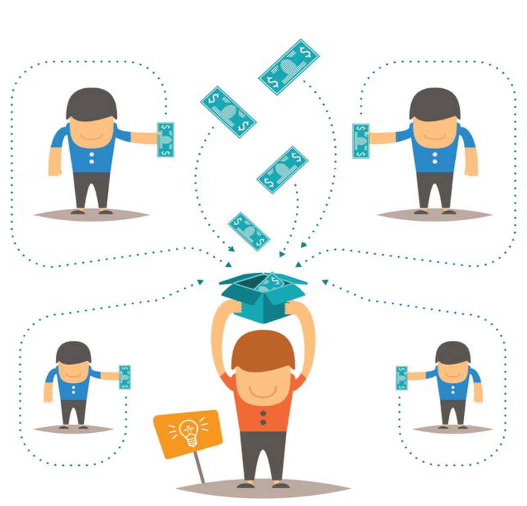
Our work investigates whether the wisdom of the lending crowd can help estimate long-term success in peer-to-peer lending. We begin by developing a set of features that characterise lending behaviour and account for potential learning effects and changes in the lender population. We then aggregate these features to describe the lender crowd contributing to individual projects. Finally, we train classification models with loan, borrower, and lending dynamics features to investigate which factors are associated with loan payment. By improving on a random-estimator baseline, we establish a basis for non-trivial prediction of loan payment from features which summarise lenders' actions and are determinants of long-term project success. We conclude by investigating the lending dynamics features that make prediction possible and further investigate whether the lending crowd is wiser in predicting the long-term success of certain project categories than others.



PROSPER MARKETPLACE

OUR DATA CONTAINS:

- 3,948,777 Bids
- 28,935 Listings
- 26,404 Borrowers
- 50,264 Lenders
- 28,935 Loans



Prosper.com is a peer-to-peer lending platform that allows borrowers to receive funding from members of a large online marketplace. Borrowers request loans by creating listings and specifying the maximum interest rate they are willing to pay if the listing turns into a loan. Lenders then bid to fund a fraction of the amount at a chosen interest rate. When a listing reaches at least 100% of its requested amount, bids with the lowest interest rates are pooled into a single loan awarded to the borrower at a final interest rate determined by Prosper.

Henry K. Dambanemuya and Emőke-Ágnes Horvát. 2019. Harnessing Collective Intelligence in P2P Lending. *In 11th ACM Conference on Web Science (WebSci '19), June 30-July 3, 2019, Boston, MA, USA.* ACM, New York, NY, USA, 9 pages.

RESEARCH QUESTION

Is the collective intelligence of lenders that emerges from online signaling better at predicting loan payment than wide-spread measures of creditworthiness like credit score?

CONTRIBUTIONS

1. *We provide new knowledge about lending dynamics* such as lenders' previous successes, the bid amount per second, as well as the coefficients of variation and herding that suggest that lenders' collective intelligence can be harnessed to improve the efficiency of crowd financing.
2. *We provide new insights about novel expressions of collective intelligence* and potential ways to harvest it. We expect that our results will inform further research into crowd-aware system design on crowd financing platforms and beyond.
3. *The proposed collective intelligence signals are general* and easily transfer to various other crowdsourcing settings.

METHODOLOGY

TO PREDICT LONG-TERM SUCCESS, WE:

1. Represent listings by a set of 23 features
2. Use standard scaling to normalise the feature set
3. Indicate loan payment (1) or default (0) in class variable
4. Tackle this binary classification task with a variety of supervised learning methods for comparison
5. Perform out of sample tests using 5-fold cross validation and show the results in Table 1
6. Compute features' Random Forest Variable Importance (viRF) score via Gini Importance as shown in Table 2

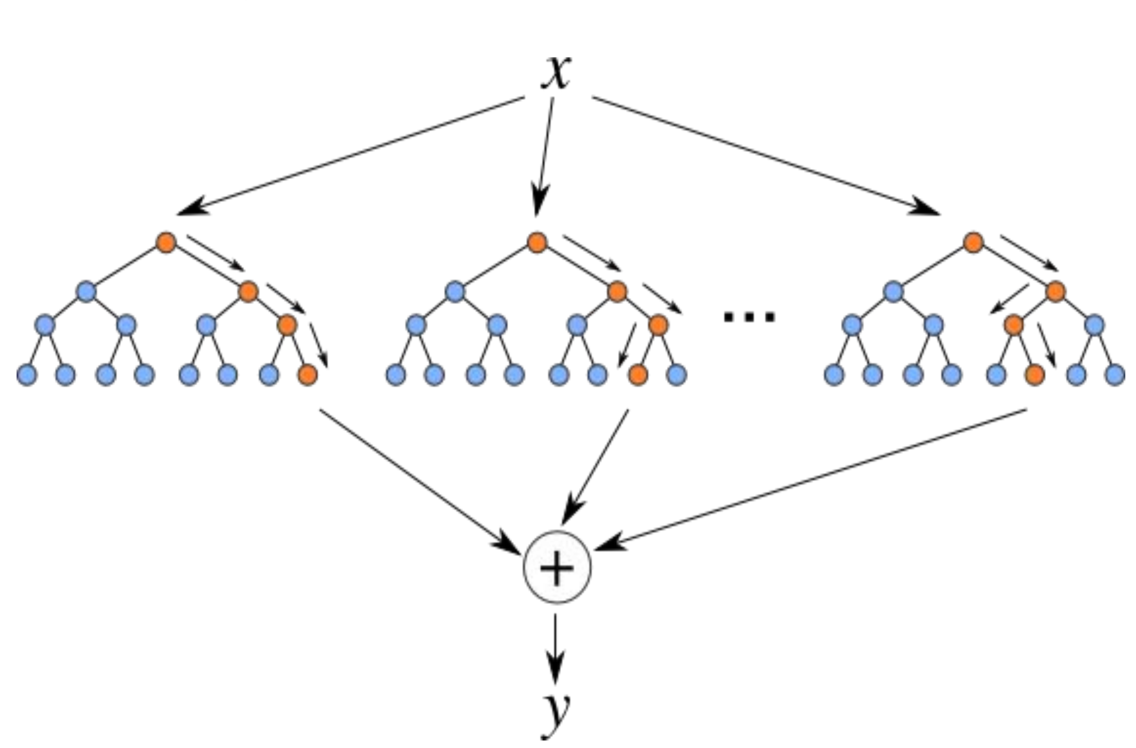


Figure 1: Illustration or Random Forest learning method for classification

RESULTS

RESULT 1: *Achieved a random forest prediction accuracy of* 0.7147. 95% confidence interval: (0.7144, 0.7149), and AUC score of 0.707.

RESULT 2: *Lenders demonstrate collective intelligence* as lending dynamics features achieve 83.7% of the predictive power of all the indicators of long-term project success combined. Additionally, lending dynamics features consistently improve estimation accuracy across loan categories as shown in Figure 2.

RESULT 3: *Lending dynamics features consistently improve estimation accuracy across loan categories* and account for 38.6% of the predictive performance.

RESULT 4: *Average lender age and experience are among the most important lending dynamics features.* Among loan categories (Figure 3) average lender experience is overall the most important.

RESULT 5: *Lending dynamics features have relatively higher viRF scores* compared to most borrower features as well as other loan features such as credit grade, requested loan amount, and loan age (Table 2).

Table 1: Long-term success prediction results. Random Forest classifier yielded best estimation results with an AUC score of 0.707.

Model	Accuracy	Precision	Recall	F-Score	AUC
QDA	0.704	0.766	0.765	0.766	0.682
CART	0.645	0.721	0.714	0.718	0.620
GNB	0.694	0.749	0.775	0.762	0.665
RF	0.715	0.797	0.736	0.765	0.707
LR	0.626	0.865	0.483	0.620	0.677
ADB	0.725	0.754	0.837	0.793	0.685

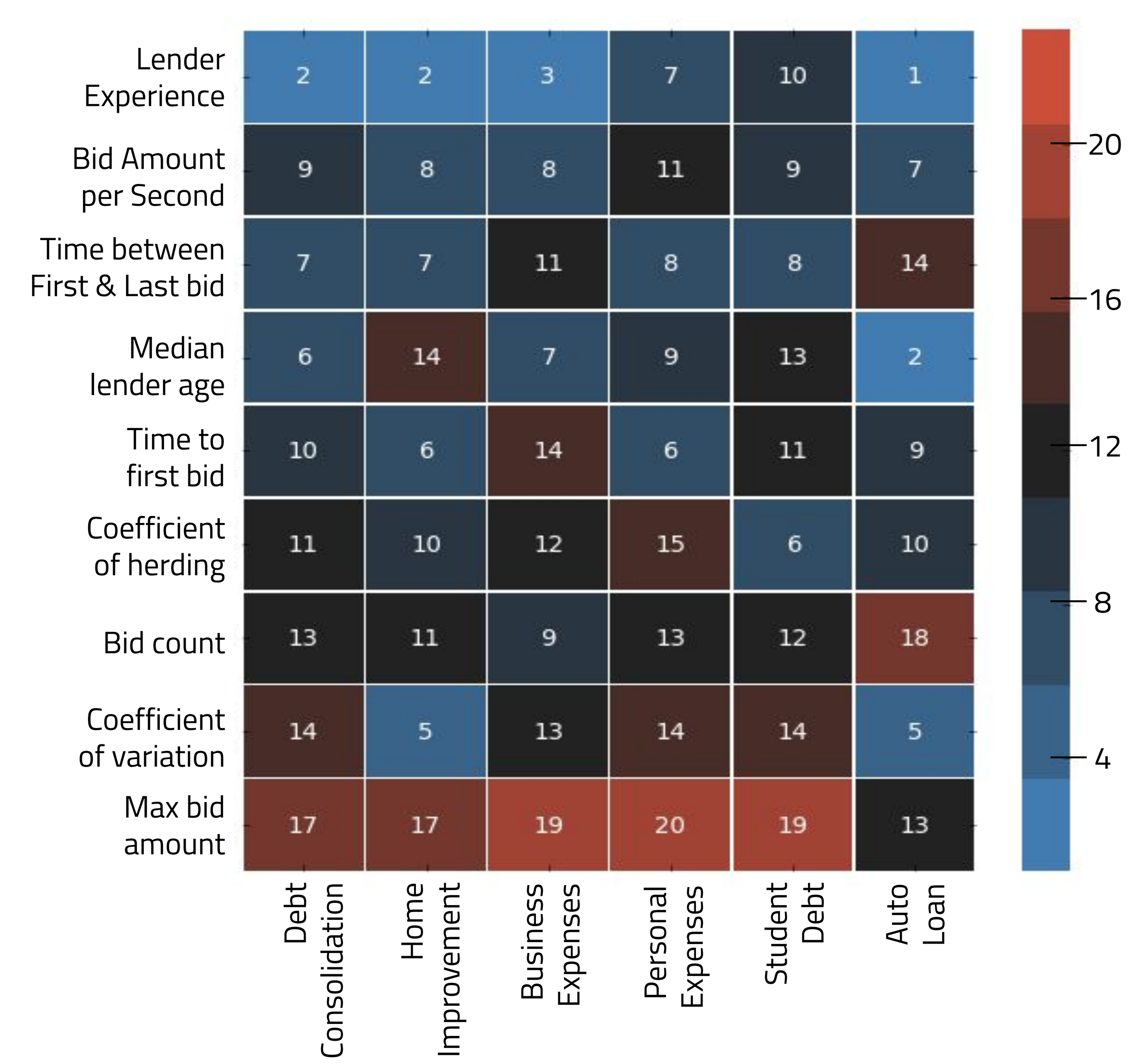


Figure 2: Analysis of feature ranking by project category. Displayed are only the ranks for lending dynamics features. We observe no systematic ranking patterns among lending features, besides the most and least predictive features: lender experience and maximum bid amount, respectively.

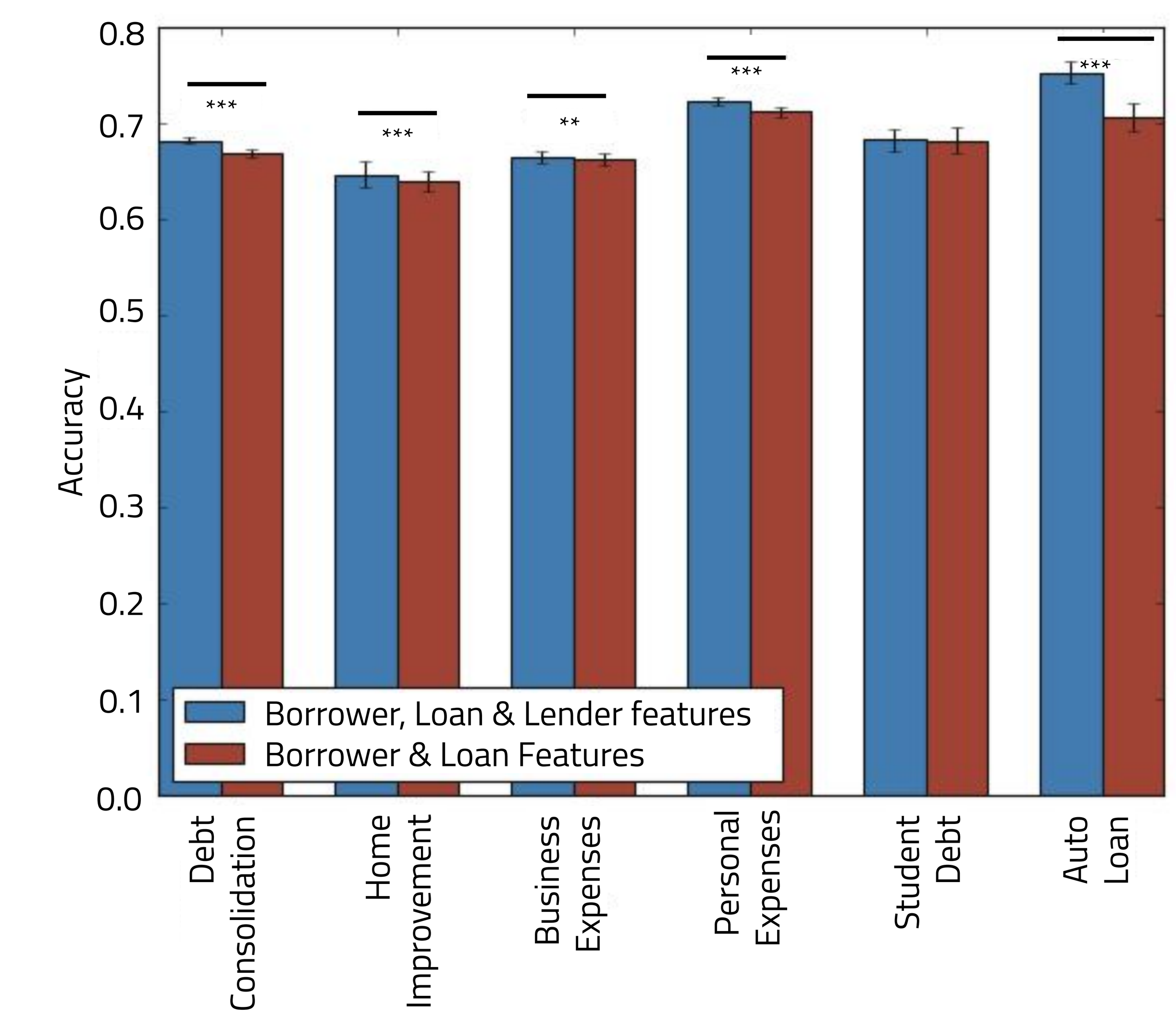


Table 2: Feature importance ranking of predictive features for long-term success categorised into groups and ranked by feature importance score.Lending dynamics features have relatively higher viRF scores compared to most borrower and loan features, accounting for 38.6% of the predictive performance. * denotes sensitive credit information.

Figure 2: Model accuracy by feature group and project category. Lender features contribute significant marginal gains in estimation accuracy indicative of collective intelligence beyond summarising loan and borrower information.

Group	Feature	Rank	viRF
Loan	Interest Rate	1	0.124
	Monthly Loan Payment	2	0.069
	Estimated Loss	3	0.055
	Description Length	8	0.045
	Credit Grade	15	0.039
	Loan Age	17	0.035
	Amount Requested	18	0.034
Lender	Average Lender Experience	5	0.052
	Bid Amount per Second	6	0.045
	Time between First and Last Bid	7	0.045
	Median Lender Age	9	0.045
	Time to First Bid	10	0.044
	Coefficient of Herding	11	0.043
	Bid Count	12	0.043
	Coefficient of variation	13	0.042
	Max Lender Bid Amount	20	0.027
Borrower	Prosper Score*	4	0.053
	Debt-to-Income Ratio*	14	0.041
	Borrower Age	16	0.038
	Credit Volatility*	19	0.028
	Role Count	21	0.025
	Borrower Experience	22	0.019
	Homeownership	23	0.010

ACKNOWLEDGEMENTS

The authors would like to thank Brian Uzzi and Jayaram Uparna for providing the data. The authors would also like to thank the anonymous referees for their valuable comments and helpful suggestions. The work is supported by the U.S. National Science Foundation under Grant No. IIS-1755873