Reputation-Aware Data Cleaning for Mobile Crowdsensing via Correlated Data

Yujian ‘Charles’ Tang, University of North Carolina, Chapel Hill, yujian@live.unc.edu
Samia Tasnim, Florida International University
Dr. Niki Pissinou, Florida International University

INTRODUCTION

- Recent increase in use of smartphones has given rise to a new, people-based data collection mode
- Many factors affect the integrity of the data (eg signal, battery, etc) [1]
- Mobile crowdsensing is useful in collecting large scale data because it operates in an open environment

MOTIVATION

- We want to ensure data accuracy but traditional security methods can’t be applied
- Huang et al. show that a reputation framework helps correct for sensing errors and protects against malicious attackers [2]
- Kang et al. show that using correlated data increases data accuracy [3]

HYPOTHESIS

Applying a reputation framework on each data type and using correlated data to predict the value of the target type will give a more accurate value than a mean-based method

METRICS*

Cooperation:

$$p = \frac{\frac{\frac{1}{\sqrt{|x_i - r|}}}{\sum_{i=1}^{N}\sqrt{|x_i - r|}}}{\sum_{j=1}^{N}\frac{\frac{1}{\sqrt{|x_j - r|}}}{\sum_{i=1}^{N}\sqrt{|x_i - r|}}}$$

Normalized Cooperation:

$$p_i^{norm} = \frac{2(p_i - min(p))}{max(p) - min(p)} - 1$$

Aged cooperation:

$$p'_{i,k} = \sum_{k'}^{k} \lambda^{k-k'} p_{i,k}^{norm}$$

Reputation (Gompertz):

$$R_{i,k}(p'_{i,k}) = ae^{be^{cp'_{i,k}}}$$

Robust Average:

$$r = \sum_{i=1}^{n} p_i * x_i$$

* i denotes the sensor, k denotes the epoch

METHODS

First discretize the space into grids and the time into epochs. Per epoch:

1. Run the Expectation Maximization algorithm (EM) with just the reputable sensors to get the first robust average and cooperation scores
2. Check if any of the disreputable sensors report accurate data and take them into account
3. Run EM again on the new set of sensors to get the new robust average and cooperation scores
4. Normalize to [-1, 1] and age cooperation scores
5. Use aged cooperation scores to calculate the new reputation of all the sensors taken into account this epoch

Algorithm 1 Expectation Maximization

Input: Robust Average (r), Cooperation Scores (p_i)
Output: Robust Average (r)

Initialize : all p_i to $1/n$ where n is the number of sensors and iteration (l) to 0
1: **while** p_i^l and p_i^{l+1} do not converge **do**
2: Compute r^{l+1} from p_i^l using Robust Average
3: Compute p_i^l from r^l using Cooperation
4: $l \leftarrow l + 1$
5: **end while**
6: **return** r^{l+1}

RESULTS

I use increase in accuracy of the prediction of the target data type using correlated data as the metric for success. Figures 1-3 show the percent improvement over a mean-based method on three different test data sets

- Figure 1 shows the framework is more accurate in 76% of cases, is 24% more precise on average, and has an average percent error of 6.8%
- Figure 2 shows the framework is more precise in 74% of cases, is 8% more precise on average, and has an average percent error of 12%
- Figure 3 shows the framework is more precise in 81% of cases, is 17% more precise on average, and has an average percent error of 9.2%

Figure 1

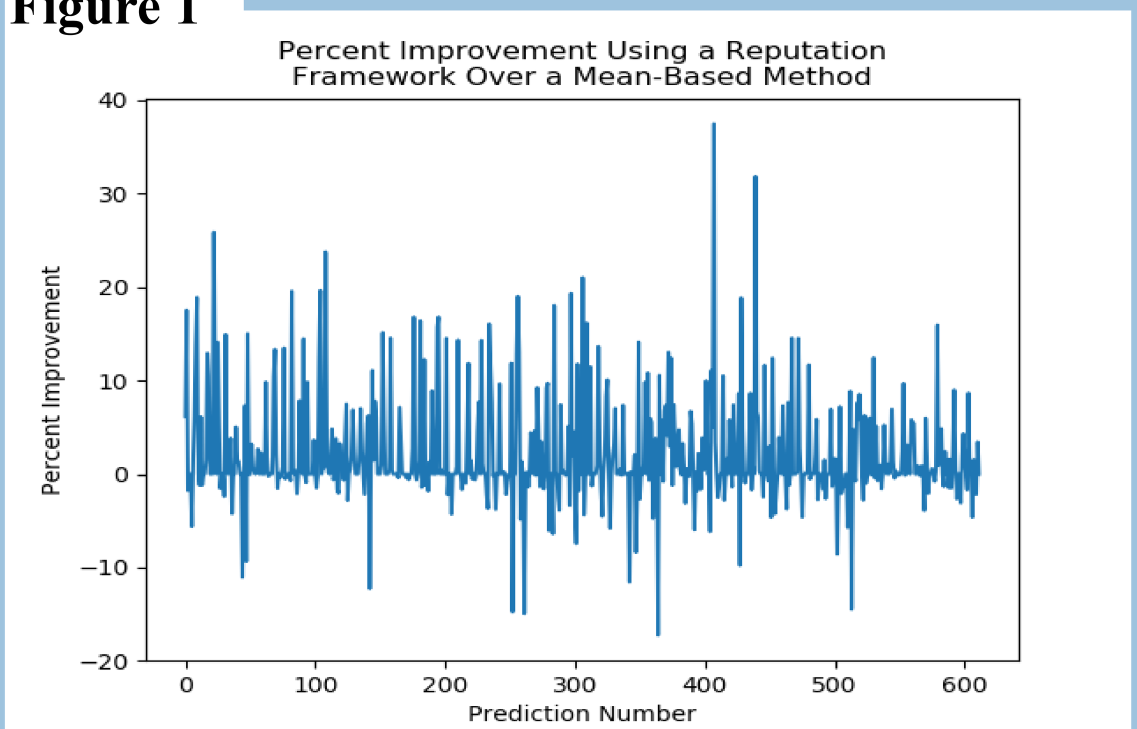


Figure 2

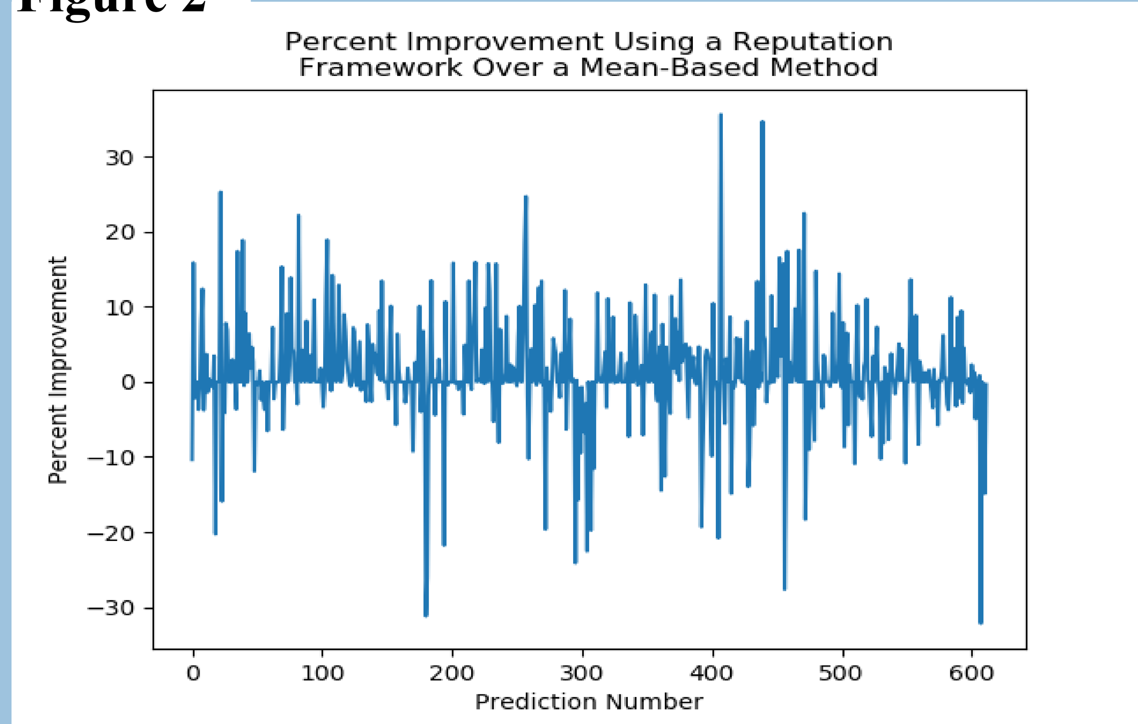
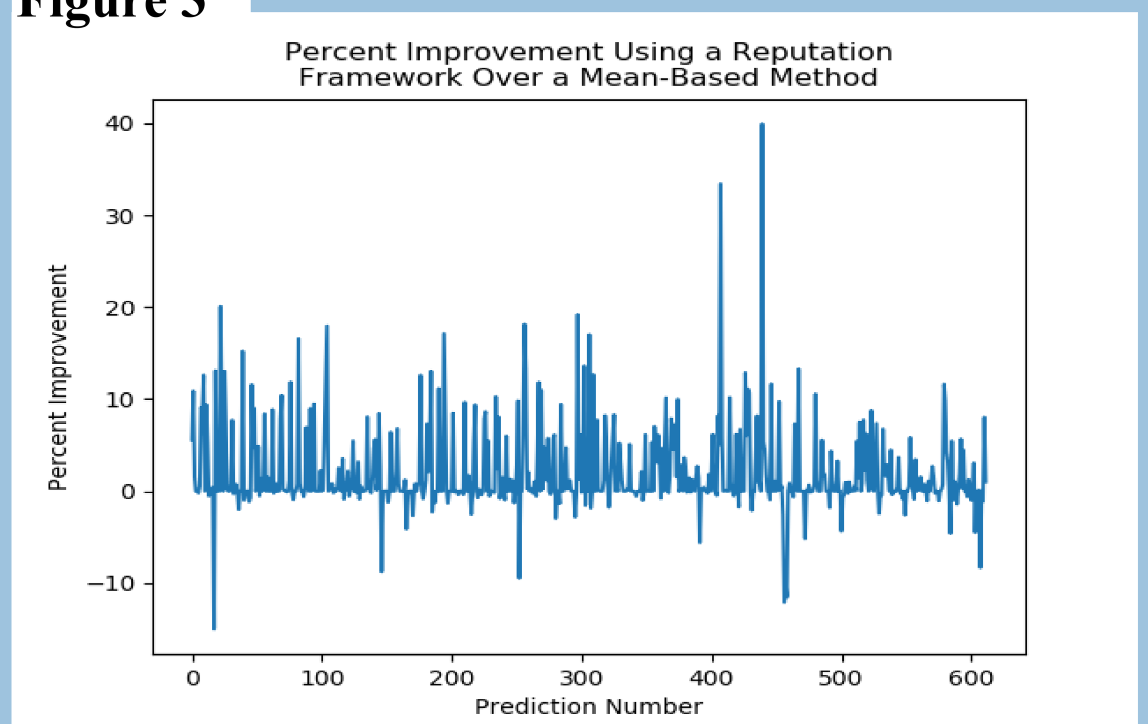


Figure 3



CONCLUSION

- In the three test cases the reputation-aware data cleaning method was more precise an average of 77% of the time, was an average of 16% more precise, and only had an average percent error of 9.3%.
- These results show that the framework has a higher accuracy overall and a high enough percentage of the time to show that it outperforms the mean-based data cleaning method
- Possible future works include using correlated data to supplement consensus-based reputation, using correlated data to fill in missing data pre-framework, and testing against other data prediction methods such as regressions

REFERENCES

1. Restuccia, Francesco, Nirnay Ghosh, Shameek Bhattacharjee, Sajal K. Das, and Tommaso Melodia. "Quality of Information in Mobile Crowdsensing: Survey and Research Challenges." *ACM Transactions on Sensor Networks (TOSN)* 13, no. 4 (2017): 34.
2. Huang, Kuan Lun, Salil S. Kanhere, and Wen Hu. "On the need for a reputation system in mobile phone based sensing." *Ad Hoc Networks* 12 (2014): 130-149.
3. Kang, Xu, Liang Liu, and Huadong Ma. "Enhance the quality of crowdsensing for fine-grained urban environment monitoring via data correlation." *Sensors* 17, no. 1 (2017): 88.

ACKNOWLEDGEMENTS

Supported by NSF Grant CNS-1560134 to the Research Experience for Undergraduates, Advanced Secured Sensor Enabling Technologies
Eric Xu