Enhancing Crowdsourced Applications via Incorporated Practice Sessions

Kinga B. Faragó, Zoltán Á. Milacski, András Lőrincz
Dept. of Software Technology and Methology,
Eötvös Lorand University,
Budapest, Hungary H-1117
{kinga.farago, srph25}@gmail.com,
lorincz@inf.elte.hu

András Németh
Synchronoss Technologies Inc.
Dublin, Republic of Ireland
anemeth@gmail.com

Abstract

In this paper, we suggest an upgraded crowdsourcing system design, where workers of the scheme can develop their skills by means of practicing in an embedded simulated environment. We show that training in the virtual setting considerably improves performance in a real traffic estimation scenario. The prototype was implemented as an Android application.

Keywords: Crowdsourcing, Traffic Estimation, Online Supervised Learning

1 Introduction

Industrial or scientific projects, which need large amount of data, can be served by non-expert crowdsourcing contributors. This is a favourable approach, since it is an economic and simple way to collect information.

However, it has a critical issue related to the quality of the provided material due to the lack of crowd proficiency. This is typically compensated by two methods: aggregating over the multiple data sources [Hun13], or incorporating specialists to supervise the accumulation [Eick14]. While these recipes have been applied successfully in many cases, the possibility of reducing the number of poor submissions is less discussed in the scientific literature.

Our goal is to settle this problem by utilizing a training process for the crowd labour, where the precision of the participants can be measured, the relevant features of the situation can be learned and applied in real scenarios.

We chose city traffic as our demonstrative example. We tested the hypothesis whether significant improvement in average speed estimations can be achieved in an actual transportation setting through a virtual training program, consisting of simulated traffic scenes with immediate feedback to the players.

2 Experimental Setup

In this section, we briefly describe the data collecting methods, the used Android applications, the training stage and the pre-processing steps of the data set.

2.1 Human Computation Task

We conducted two outdoor experiments in a real traffic environment: one before virtual training and one afterwards. In both trials, 4 test subjects were asked to travel on the same tramline (in downtown Budapest), and report estimated average speed values of their surroundings with our crowdsourcing Android application. Meanwhile, a driver was also asked to travel around the same route by car, and track the ground truth momentary speed frequently by an automatic GPS logging mobile software. Together with the momentum information, the two programs have also stored the exact GPS coordinates and Unix time stamps of all measurements. Figure 1 portrays screenshots of the user interfaces.

Instead of an implemented virtual training session, a mock-up procedure was held between the two outdoor experiments. It consisted of watching a series of 47 videos created in the Vissim (http://www.ptvgroup.com) traffic simulator. Figure 2 shows excerpts of the

In: F. Hopfgartner, G. Kazai, U. Kruschwitz, and M. Meder (eds.): Proceedings of the GamifIR'15 Workshop, Vienna, Austria, 29-March-2015, published at http://ceur-ws.org

recordings. The videos were set in the exact same path as in the live tasks, but in a virtual environment (either on the street or on the identical tramline). Many kinds of travelling scenarios were generated, ranging from jams to completely free roads, to adapt for different events. Each video segment was 1 minute long and it was adjusted to the traffic light switching cycle. Subjects were asked to estimate the average velocities of the vehicles in an online supervised learning [Sut93] manner: the true speed was given at the end of the videos. Subjects were expected to leverage their newly acquired skills in the second outdoor experiment. Note that 2D screens are suitable for training, since one can deduce 3D scenes by structure-from-motion mechanisms [Tre91].

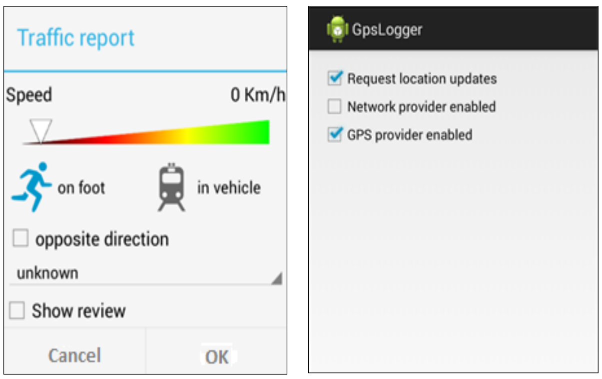


Figure 1: Screenshots of the Crowdsourcing (left) and GPS Logging (right) Android Applications

2.2 Data Set

In the first experiment, 3129 car and 128, 198, 181, 141 human samples were reported. In the second experiment, 1736 automobile log records and 77, 115, 107, 77 manual measurements were registered, respectively. From each human data set, 30 elements were carefully selected for further processing by minimizing the spatio-temporal distances from the car. Together with this operation, we discarded vehicle records that have become unnecessary, keeping only 601 and 676 points, as can be seen on Figure 3.

Three extra pre-processing steps were imposed on the data in Matlab R2013b. Firstly, the car measurements were not uniformly sampled in time. Linear interpolation was used to overcome this matter. Secondly, the car speed values were momentary, thus we had to obtain averages. This was carried out by the Fused LASSO algorithm [Tib05], which is capable of detecting segments with constant mean within a time series. Finally, the human and the car reports were not exactly aligned with each other in space and time, so we had to match the two data sets. With the Dynamic Time Warping procedure [Sak70], we coupled each human observation with a corresponding car datum. Table 1 illustrates the obtained pairs of true and estimated average speeds.

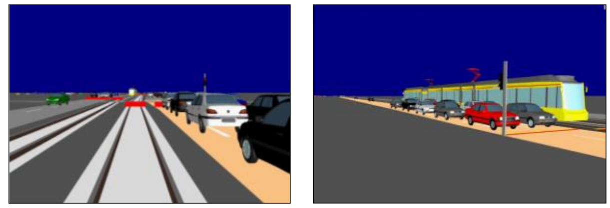


Figure 2: Screenshots of the Simulated Training Environment on (left) and beside (right) the Tramline

3 Experimental Results

To compare the results from the two outdoor experiments, we ran hypothesis tests on the pre-processed data sets in Matlab R2013b. Squared differences between the coupled ground truth car speeds and human estimations were computed for additional analysis. Since these values were not normally distributed, we conducted nonparametric one-sided Wilcoxon rank sum tests for paired deviation data sets of each subject (at level α=0.05). Obtained p-values for the participants were 0.01460, 0.00515, 0.10578 and 0.02971, respectively. The tests confirmed that the median error values were significantly smaller in the second experiment for subjects 1, 2 and 4. Improvement was not considerable for subject 3. Corresponding root-mean-square deviation (RMSD) performance metrics can be found in Table 1.

4 Conclusion and Future Research

We found a possibility for significant improvement in human traffic estimation between two outdoor crowdsourcing experiments. Test subjects were asked to estimate average transportation speeds of their surroundings. The first attempts were performed without any prior experience in this task, the second trials were carried out after completing training exercises in a simulated version of the same location.

This novel result shows that integrating experience gaining processes within crowdsourcing applications can lead to higher quality submissions in real scenarios. Our findings are, however, limited by the small number of test subjects and the lack of a control group. Motivating participants to complete such training tasks is also a question that we did not consider here: gamification would be the proper way to examine this. Further studies could be carried out to address these issues.

We demonstrated that crowdsourcing training exercises can pave a way for more efficient data services and – consequently – better user experiences.

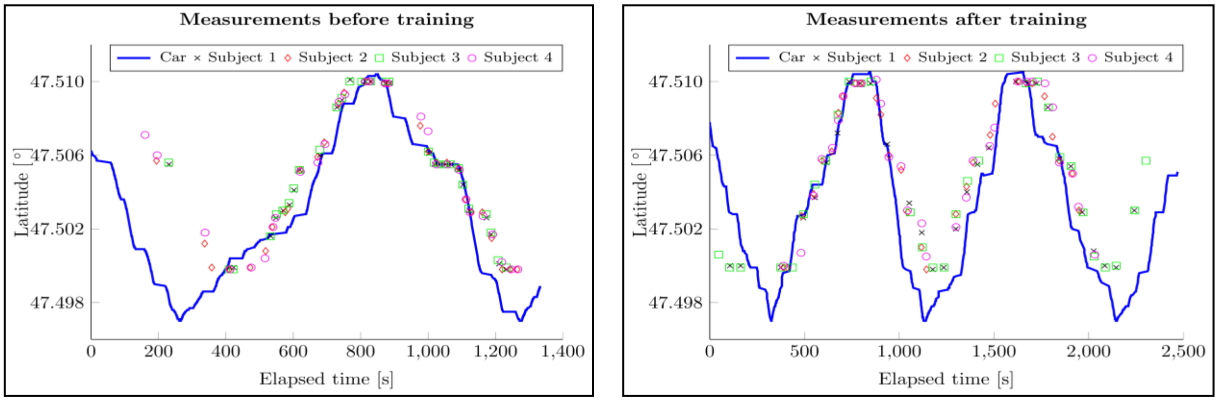
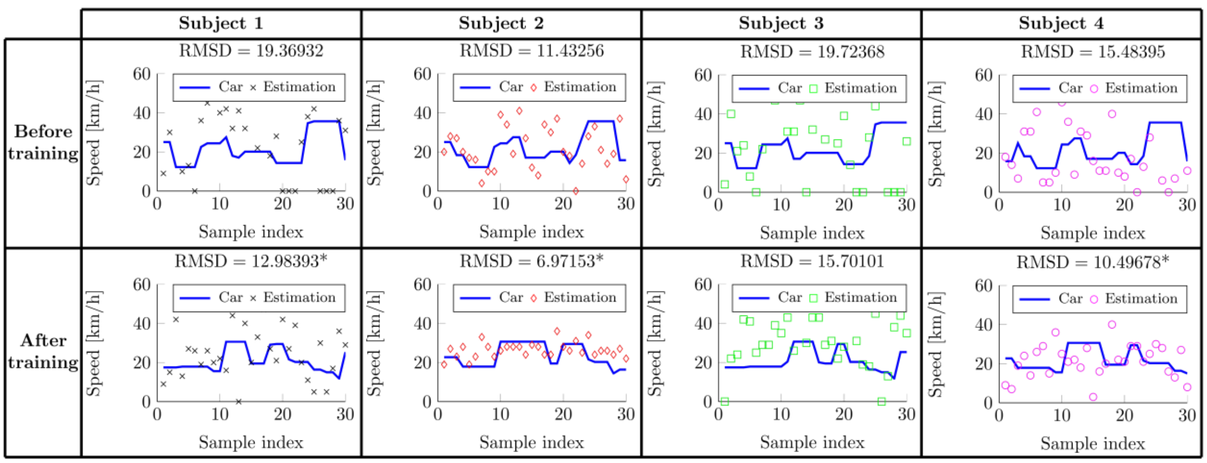


Figure 3: Chosen Data Points before (left) and after (right) Training

Table 1: RMSD Metric of Matched Human and Car Speed Measurements (*=statistically significant improvement)



Acknowledgements

We would like to thank our participants for their cooperation and persistent work. We are also grateful to Tamás Tettamanti for the Vissim simulations. The research was performed in cooperation with the EIT ICT Labs Budapest Associate Partner Group. (www.ictlabs.elte.hu)

References

[Eick14] Eickhoff, C. Crowd-powered experts: Helping surgeons interpret breast cancer images. In *Proc. of the 1st Int. Workshop on Gamification for Information Retrieval*, (Amsterdam, NL, 2014), ACM, 53-56.

[Hun13] Hung, N. Q. V., Tam, N. T., Tran, L. N., and Aberer, K. An evaluation of aggregation techniques in crowdsourcing. In *Proc. of the 14th Int. Conf. on Web Information Systems Engineering*, (Nanjing, CN, 2013), Springer Berlin Heidelberg, 1-15.

[Sak70] Sakoe, H., and Chiba, S. A similarity evaluation of speech patterns by dynamic programming. In *Proc. of the Nat. Meeting of Inst. of Electronic Communications Engineers of Japan*, (JP, 1970), 136.

[Sut93] Sutton, R. S., and Whitehead, S. D. Online learning with random representations. In *Proc. of the 10th Int. Conf. on Machine Learning*, (Amherst, US, 1993), 314-321.

[Tib05] Tibshirani, R., Saunders, M., Rosset, S., Zhu, J., and Knight, K. Sparsity and smoothness via the fused lasso. *J. of the Royal Statistical Soc. Series B (Statistical Methodology)*, *67* (1), (2005), Wiley, 91-108.

[Tre91] Treue, S., Husain, M., & Andersen, R. A. Human perception of structure from motion. *Vision Research*, *31* (1), (1991), Pergamon Press, 59-75.