Communication-efficient learning of traffic flow in a network of wireless presence sensors

Marco Stolpe, Thomas Liebig, and Katharina Morik

TU Dortmund, Computer Science, LS 8, 44221 Dortmund, Germany

Abstract. Current traffic management systems learn global traffic flow models based on measurements from a static mesh of hard-wired presence sensors. The centralization of all data comes at the cost of limited scalability, security and fault-tolerance. Modern traffic control could benefit from a decentralized system of cheap wireless sensors. However, constrained devices pose challenges for data analysis, which must be communication- and energy-efficient as well as secure. We hereby present a privacy-preserving decentralized in-network algorithm which exchanges space-time aggregated values between restricted sets of topological neighboring nodes. The algorithm's evaluation on real world traffic data demonstrates its performance in terms of communication cost and accuracy.

Keywords: vertically distributed, communication costs, label proportions, spatio-temporal models, traffic prediction

1 Introduction

Traffic flow prediction allows intelligent control of traffic lights and other traffic signals. Individual mobility benefits from predictions as well, since they allow for proactive, smart decisions on individual travel plans like the avoidance of likely traffic hazards [9, 4]. Centralization of all data and control can easily account for global traffic patterns and relationships. However, a single point of failure poses high security risks facing natural disasters or intended devastation. The server-side collection further may become a bottleneck for real-time processing and is thus not scalable. The maintenance of cable networks is costly regarding materials and construction work. Moreover, the area of traffic management systems is often limited by the political area of homogeneous network regulations.

Therefore, we propose a decentralized system of cheap battery-powered wireless presence sensors that work mostly autonomously and may easily be attached to existing infrastructure like traffic lights, signs or buildings, increasing coverage. Moreover, restricting learning to topologically close sensors could make traffic control more robust facing disasters, as failures affect only confined parts of the whole system. However, decentralized wireless networks pose their own challenges by putting severe constraints on data analysis tasks. These involve questions of streaming data, dynamic network changes, and concept drift. The work at-hand extends our previous paper [5]. In this paper, we restrict ourselves to the following questions:

– The energy and bandwidth of small battery-powered devices are highly limited. How can we reduce the amount of data communicated and what are trade-offs in terms of accuracy?
– Increasing coverage and the network's density bares the risk of identifying individual persons and tracking them. How can we guarantee their privacy?

We present a distributed spatio-temporal in-network learning algorithm that exchanges only space-time aggregated values between topologically close sensors, reducing communication and providing k-anonymity by design. The algorithm has been evaluated on real world traffic data from the city of Dublin, where the focus is on the prediction of future traffic flow at junctions throughout the city.

2 Related Work

Distributed algorithms mostly focus on horizontally partitioned data, whereas our data is vertically distributed. Here, privacy-preserving SVMs like [18] are not scalable, as they send and process quadratic kernel matrices. Distributed optimization algorithms like [1] iteratively exchange predictions for each observation, potentially sending more than the entire dataset. Anomaly detection algorithms, like [14], require a central coordinator. [3] trains local SVM models, but labels are sent by a central coordinator.

Algorithms for learning traffic flow mostly work centrally as well. Simulations [12] and imputation models [6] estimate traffic flow at unobserved locations, while our study focuses on predictions at sensor locations. The Kriging approach [16] is based on spatial relations, like [7], whose approach gives measured segments close to unmeasured ones a higher impact. We utilize this idea to build local models based on their closest sensors. More complex approaches investigate neural networks or SVMs. A Gaussian Markov Model was proposed in [13], and Spatio-Temporal-Random-Fields (STRFs) in [11]. The few distributed approaches count and re-identify individual vehicles, while we use aggregated quantities.

The task of learning from aggregated label information was first introduced in [2]. Theoretical bounds have recently been proven in [17]. For further references to actual learning algorithms, see [10]. In this paper, we adapt LLP [15], that due to its linear running time and small memory footprint fits well to a constraint scenario.

3 Distributed Learning of Spatio-Temporal Local Models

Given are m wireless sensor nodes $i = 1, \dots, m$. Each node i delivers an infinite series of real-valued raw measurements $\dots, v^{t-1}(i), v^t(i)$, $v^{t+1}(i), \dots$, where t denotes the current time step t and $t-1/t+1$ denote next and previous ones, assuming a constant sample rate. Associated with each sensor is a spatial location. Decisions on traffic signals often base on discrete flow categories, achieved by a mapping $d : \mathbb{R} \to Y$ of raw measurements to categories $Y = \{Y_1, \dots, Y_l\}$.

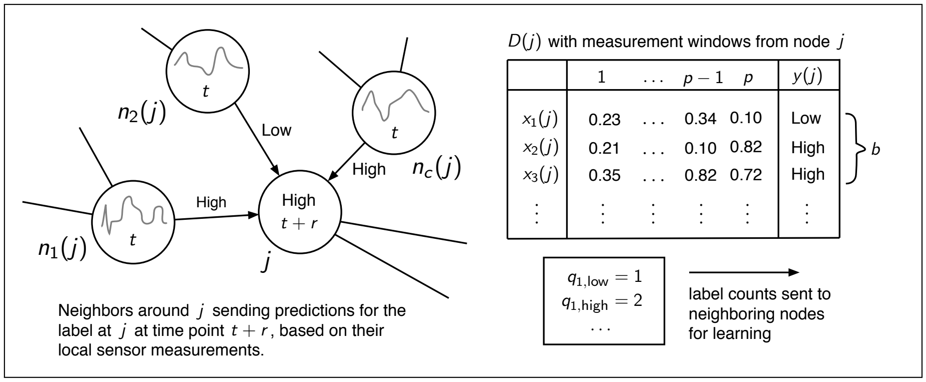


Fig. 1: Distributed Learning of Local Models

Let each node j provide a set $D(j)$ for supervised offline learning, containing n pairs $(x_i(j), y_i(j))$ of training examples $x_i(j) \in \mathbb{R}^p$ and labels $y_i(j) \in Y$. Each $x_i(j)$ is created by sliding a window of size p with step size 1 over the stream of measurements at node j. When recording from time step s, observations $x_i(j) = [v^{s+i-1}(j), \dots, v^{s+i-2+p}(j)]$ are windows of measurements and labels $y_i = d(v^{s+i-2+p+r}(j))$ are discretized measurements r time steps ahead.

For every node j, we restrict learning to j and c neighboring nodes with indices $n_1(j), \dots, n_c(j)$. Based on the datasets at j and its neighbors, we want to learn a *local* model $f(j)$ that, given current windows $x(j), x(n_1(j)), \dots, x(n_c(j))$ of sensor readings, predicts the traffic flow category $y(j)$ at node j with horizon r correctly. This situation is depicted in Fig. 1, showing node j with its neighbors and an exemplary dataset. Interpreting measurement windows as features of a single observation x, the data is *vertically partitioned*, since each neighboring node of j only stores partial information about x, i.e. a subset of features.

One way to learn is to send measurements from neighbors to j, combine the according windows with j's labels and, based on this data, learn $f(j)$ at j. Another way is to combine windowed measurements at each neighbor with labels from j, i.e. $D_j(k) = \{(x_i(k), y_i(j))\}_{i=1,\dots,n}$, and learn models $f_j(k)$ at nodes $k = j, n_1(j), \dots, n_c(j)$ to predict $y(j)$. Model $f(j)$ could then be a majority vote over predictions from j and its neighbors. The first approach may respect joint dependencies between nodes, but isn't privacy-preserving, because it sends raw measurements. We opt for the latter approach, since a discretization of values preserves privacy and saves communication by encoding the data in less bits.

However, privacy and communication can be improved even further. Given a partitioning of observations $x_1, \dots, x_n$ into batches B_u, $u = 1, \dots, h$ and label proportions π_{uv} for each batch u and class Y_v, algorithms for *learning from label proportions* (see Sect. 2) learn model $f : X \to Y$ that assigns labels to individual observations. Instead of sending all labels to neighbors, it might therefore suffice to send only the counts of labels per batch.

A simple partitioning of the data into b-sized batches is a division over consecutive time intervals. Node j counts how often each class occurs in each batch and sends these counts in a $h \times l$ matrix $Q(j)$, $h = \lceil n/b \rceil$, to its neighboring nodes. These transform $Q(j)$ into a label proportion matrix $\Pi(j)$, yielding the original problem of learning from label proportions at neighboring nodes of j.

In principle, arbitrary learners for the problem could be used. We have adapted the LLP algorithm [15], since it fits to our constraint scenario and can handle multiple classes. LLP first clusters all observations ($k \geq |Y|$) and then minimizes the mean squared error (MSE) between the given label proportions and those as calculated by different label assignments to clusters. We abstain from the evolutionary optimization of attribute weights as described in the original paper. Further, we have replaced the exhaustive labeling strategy by a more efficient local search with a multistart strategy, which still yields a sufficient accuracy. This modified version will be called LLP_{ms}. The distributed learning algorithm after preprocessing now works as follows:

For $j = 1$ **to** m **do** /* *in parallel* */
  divide $D(j)$ into batches $B_1, \ldots, B_h$
  calculate label counts for each batch and store them in $Q(j)$
  send $Q(j)$ to nodes $n_1(j), \ldots, n_c(j)$
  For $k = j, n_1(j), \ldots, n_c(j)$ **do** /* *in parallel* */
    calculate $\Pi(j)$ from $Q(j)$
    train LLP_{ms} model $f_j(k)$ at node k

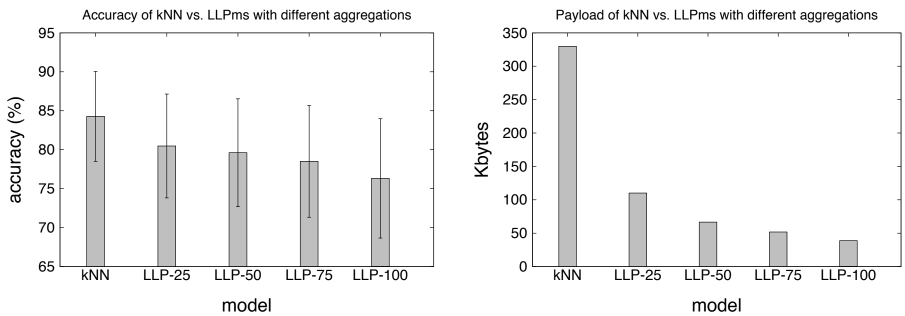
Each node stores $c+1$ different models, for itself and each of its neighbors. All models are *local* in the sense that learning is restricted to local neighborhoods. Moreover, the algorithm works fully *in-network*, as no central coordinator is needed for local synchronization and learning between peer nodes.

4 Experiments

Methods are evaluated on traffic flow data at junctions in the city of Dublin, recorded by the Sydney Co-ordinated Adaptive Traffic System (SCATS)[1] [8]. For January 2013, we average measurements at arms of a junction and aggregate them over 15 minute intervals, resulting in 2,976 time slices at 296 sensor locations. For learning traffic flow categories at the next 15 minutes, based on previous time slices, we create datasets $D(j)$ by sliding a window ($p = 5$, 75 minutes) over node j's measurements. Labels are discretized traffic flow values at horizon $r = 1$ according to ranges 0-5, 5-30, 30-60, 60-150 and 150-260.

LLP_{ms} is trained at j and six nearest nodes (according to Euclidean distance), based on j's label counts for batch sizes $b = 25, 50, 75$ and 100. As baseline, we also use kNN ($k = 15$) with all labels and predictions from a global STRF model [11]. As performance we take the average accuracy over all nodes, assessed by a 10-fold cross validation per node.

[1] Data is publicly available at http://dublinked.ie .



(a) Accuracy (b) Payload

Fig. 2: Trade-off between accuracy and payload sent for kNN and $\mathrm{LLP_{ms}}$

The STRF outperforms kNN on the category of lowest traffic flow (83.3% vs. 43.3% recall), while kNN outperforms the STRF when predicting the highest traffic flow (77.3% vs. 70.5% recall). On average, however, kNN models outperform the STRF, as they achieve an average accuracy of 85.7% over all nodes, whereas the STRF yields an accuracy of only 78.1%.

Figure 2 shows the trade-off between accuracy and payload sent for kNN and $\mathrm{LLP_{ms}}$ trained on differently sized batches of aggregated labels. $\mathrm{LLP_{ms}}$'s accuracy decreases the more we aggregate, however, also the communication costs decrease, by factors 3, 5 and 8.5 in comparison to sending each individual label. For $b = 75$, the accuracy is still in the order achieved by the more complex STRF model, though much less data needs to be communicated for learning.

5 Conclusions

In this paper, we have presented a privacy-preserving and communication-efficient distributed in-network algorithm for spatio-temporal learning, together with an evaluation of its accuracy and communication costs. Though the results look promising, further questions need to be answered like that of streaming data and concept drift. Moreover, the algorithm's sensitivity to changing the number of neighbors or other parameters needs to be investigated. Finally, we want to evaluate the approach in the context of other applications.

Acknowledgements This research has received funding from the European Union's Seventh Framework Programme under grant agreement number FP7-318225, INSIGHT. Additionally, this work has been supported by Deutsche Forschungsgemeinschaft (DFG) within the Collaborative Research Center SFB 876, project B3. We thank Jan Czogalla for data preprocessing and Nico Piatkowski for carrying out the STRF experiments and fruitful discussion.

References

1. Bellet, A., Liang, Y., Garakani, A.B., Balcan, M.F., Sha, F.: Distributed Frank-Wolfe algorithm: A unified framework for communication-efficient sparse learning. CoRR abs/1404.2644 (2014)
2. Kück, H., de Freitas, N.: Learning to classify individuals based on group statistics. In: Proc. of UAI. pp. 332–339 (2005)
3. Lee, S., Stolpe, M., Morik, K.: Separable approximate optimization of support vector machines for distributed sensing. In: Proc. of ECML/PKDD. pp. 387–402. Springer (2012)
4. Liebig, T., Piatkowski, N., Bockermann, C., Morik, K.: Predictive trip planning - smart routing in smart cities. In: Proc. of EDBT/ICDT Workshops. vol. 1133, pp. 331–338. CEUR-WS.org (2014)
5. Liebig, T., Stolpe, M., Morik, K.: Distributed traffic flow prediction with label proportions: From in-network towards high performance computation with mpi. In: Proc. of MUD2, vol. 1392, pp. 36–43. CEUR-WS (2015)
6. Liebig, T., Xu, Z., May, M., Wrobel, S.: Pedestrian quantity estimation with trajectory patterns. In: Proc. of ECML/PKDD, pp. 629–643. Springer (2012)
7. May, M., Hecker, D., Körner, C., Scheider, S., Schulz, D.: A vector-geometry based spatial kNN-algorithm for traffic frequency predictions. In: Proc. of ICDM Workshops. pp. 442–447. IEEE Computer Society (2008)
8. McCann, B.: A review of SCATS operation and deployment in Dublin. In: Proc. of JCT Traffic Signal Symposium & Exhibition. JCT Consulting Ltd (2014)
9. Niu, X., Zhu, Y., Cao, Q., Zhang, X., Xie, W., Zheng, K.: An online-traffic-prediction based route finding mechanism for smart city. Int. Journal of Distributed Sensor Networks 501 (2015)
10. Patrini, G., Nock, R., Caetano, T., Rivera, P.: (Almost) no label no cry. In: NIPS 27, pp. 190–198. Curran Associates, Inc. (2014)
11. Piatkowski, N., Lee, S., Morik, K.: Spatio-temporal random fields: compressible representation and distributed estimation. Machine Learning 93(1), 115–139 (2013)
12. Raney, B., Nagel, K.: An improved framework for large-scale multi-agent simulations of travel behavior. Towards better performing European Transportation Systems pp. 305–347 (2006)
13. Schnitzler, F., Liebig, T., Mannor, S., Morik, K.: Combining a Gauss-Markov model and Gaussian process for traffic prediction in Dublin city center. In: Proc. of EDBT/ICDT. vol. 1133, pp. 373–374. CEUR-WS.org (2014)
14. Stolpe, M., Bhaduri, K., Das, K., Morik, K.: Anomaly detection in vertically partitioned data by distributed Core Vector Machines. In: Proc. of ECML/PKDD. pp. 321–336. Springer (2013)
15. Stolpe, M., Morik, K.: Learning from label proportions by optimizing cluster model selection. In: Proc. of ECML/PKDD, pp. 349–364. Springer (2011)
16. Wang, X., Kockelmann, K.M.: Forecasting network data: Spatial interpolation of traffic counts from texas data. Journal of the Transportation Research Board 2105(13), 100–108 (2009)
17. Yu, F.X., Kumar, S., Jebara, T., Chang, S.: On learning with label proportions. CoRR abs/1402.5902 (2014)
18. Yunhong, H., Liang, F., Guoping, H.: Privacy-preserving SVM classification on vertically partitioned data without secure multi-party computation. In: Proc. of ICNC. vol. 1, pp. 543–546 (Aug 2009)