The Thirty-Fourth AAAI Conference on Artificial Intelligence (AAAI-20)

Travel Time Prediction on Un-Monitored Roads: A Spatial Factorization Machine Based Approach (Student Abstract)

Lile Li, Wei Liu
School of Computer Science, Faculty of Engineering and IT
University of Technology Sydney, Australia
Lile.Li@student.uts.edu.au, Wei.Liu@uts.edu.au

Abstract

Real-time traffic monitoring is one of the most important factors for route planning and estimated time of arrival (ETA). Many major roads in large cities are installed with live traffic monitoring systems, inferring the current traffic congestion status and ETAs to other locations. However, there are also many other roads, especially small roads and paths, that are not monitored. Yet, live traffic status on such un-monitored small roads can play a non-negligible role in personalized route planning and re-routing when road incident happens. How to estimate the traffic status on such un-monitored roads is thus a valuable problem to be addressed. In this paper, we propose a model called Spatial Factorization Machines (SFM) to address this problem. A major advantage of the SFM model is that it incorporates physical distances and structures of road networks into the estimation of traffic status on un-monitored roads. Our experiments on real world traffic data demonstrate that the SFM model significantly outperforms other existing models on ETA of un-monitored roads.

Introduction

Estimated time of arrival (ETA) is an important factor for effective route and travel plannings. An accurate ETA relies on not only historical travel times but also real-time live traffic situations. Particularly, the live traffic situations are vitally important to route re-planning and detouring when incidents such as unexpected traffic congestions emerge (Du et al. 2018).

However, in most cases traffic authorities can only have the resources and budget to monitor highways and major roads. Yet, the un-monitored smaller roads, if their traffic status are accurately estimated, can provided very helpful information for alternative and complementary route planning when incidents happens (Huang et al. 2018). The estimate of traffic status for un-monitored road is nonetheless not a well-studied research area. In (Bellini et al. 2018), a few fixed traffic sensors were deployed in the areas of un-monitored roads which were then used to generate traffic estimations. This approach however faces physical constraints since it is not always feasible to install sensors on all un-monitored small roads.

In this paper, we propose the first research that uses factorization machines to estimate traffic situation on un-monitored roads. Our model does not need the installation of extra sensors, but instead can use information from monitored roads and road networks to make effective estimations of traffics on un-monitored roads. Factorization machines (FMs) (Rendle 2010) and Coupled Factorization Machines (CFMs) (Li, Do, and Liu 2019) were originally proposed for recommender systems. In this research, we improve the existing literature by proposing a Spatial Factorization Machines (SFM) which specifically takes spatial information into the training of FMs. In summary, the contribution we make in this poster paper are as follows:

1. We propose a Spatial Factorization Machine (SFM) that takes geo-graphical information into the training and optimization processes. The SFM makes it possible to use distance matrices, Laplacian matrices, and other spatial network information to guide the training of FMs.
2. We propose to use the theory of SFM to solve the important problem of traffic estimation on un-monitored roads. We perform experiments on real-world data which validate the superiority of the SFM model over other alternative approaches.

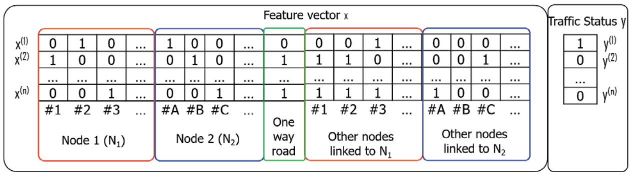
Description of Our Approach

Factorization Machines (Rendle 2010) deal with data features combined in a vector format $\mathbf{x}$. The relation between the features and the predicted label $\hat{f}(\mathbf{x})$ is modeled by 2-way interactions of the features as

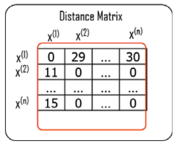
$$\hat{f}(\mathbf{x}) = w_0 + \sum_{i=1}^{n} w_i x_i + \sum_{i=1}^{n} \sum_{j=i+1}^{n} w_{i,j} x_i x_j$$

where n is the size of the feature vector, w_0 is a bias, w_i captures the strength of i-th feature and $w_{i,j}$ models the interaction between the i-th and j-th features. Moreover, the FM models the feature interactions by factorizing them: $w_{i,j} =< \mathbf{v}_i, \mathbf{v}_j >_1^k = \sum_{f=1}^{k} v_{i,f} \times v_{j,f}$, where k is the rank of the factorization and $\mathbf{v}_i$ and $\mathbf{v}_j$ are the latent vectors of the i-th and the j-th features of $\mathbf{x}$, respectively.

Suppose that a feature vector $\mathbf{x}$ contains information of a road segment as illustrated in Fig.1(a), also suppose that



(a) Road segment feature information



(b) Distance matrix among road segments

Figure 1: Feature information and distance matrix among road segments. In subfigure (b), the demonstrated distance matrix is not symmetric because many roads in real world are one-way roads.

we have a distance-based adjacency matrix as illustrated in Fig.1(b), where the distances among road segments are asymmetric (reflecting the real-world situation that many roads are one-way roads), then the underlying problem is how we can use the asymmetrical distance matrix to estimate the traffic status (i.e., being congested or not congested) of un-monitored road segments among all the road segments.

Denote the road segment information matrix (Fig.1(a)) by X and the distance matrix (Fig.1(b)) by D, our Spatial Factorization Machine (SFM) models the coupled row-wise fields between X and D through factorizing common latent vectors $\mathbf{v}$:

$$\hat{f}(\mathbf{x}, \mathbf{d})_{x \in X, d \in D} = w_0 + \sum_i w_i x_i + w_0' + \sum_i w_i' d_i \\ + \sum_i \sum_j < \mathbf{v}_i, \mathbf{v}_j >_1^k x_i x_j + \sum_i \sum_j < \mathbf{v}_i, \mathbf{v}_j' >_1^k d_i d_j$$

where k is the length of the latent factors $\mathbf{v}_i$, $\mathbf{v}_j$, and $\mathbf{v}_j'$. Similar to FM (Rendle 2010), parameters (w_0, $\mathbf{w}$, $\mathbf{v}$, w_0', $\mathbf{w}'$ and $\mathbf{v}'$) of SFM can be effectively learned by gradient descent methods, such as stochastic gradient descent.

Performance Evaluation

We evaluate the performance of SFM against other existing models using real travel time data obtained from an Australian Traffic Management Centre that transferred taxi movement data into travel times. The data covers taxi movements for most roads in the Greater Sydney regions from 1 April 2018 to 30 May 2018.

We built the distance matrix via computing the distances between all road segments using Google maps. For every road segment, we binarize the travel time into labels of *congested* and *not congested* where *congested* means the travel time is longer than the 75th percentile of all travel times of that road segment. We randomly selected 20% of the data which are used as un-monitored roads - i.e., we removed them from the training and only use them in testing. We repeat the random selection of the 20% data for 5 times and report the mean of the 5 runs.

Table 1: Comparison among different models in traffic estimation accuracy. The last row presents p-values of t-tests between each algorithm and our proposed SFM.

k	Month	FM	FFM	CFM	Our SFM
5	April	0.903	0.896	0.915	0.938
	May	0.858	0.886	0.882	0.902
10	April	0.852	0.854	0.936	0.946
	May	0.866	0.898	0.882	0.915
15	April	0.869	0.889	0.895	0.944
	May	0.874	0.872	0.882	0.937
20	April	0.888	0.864	0.903	0.935
	May	0.858	0.856	0.928	0.957
25	April	0.867	0.879	0.898	0.947
	May	0.863	0.875	0.936	0.952
p-values		2×10^{-5}	5×10^{-5}	1×10^{-4}	–

We compare our SFM model with other approaches including the FM(Rendle 2010), FFM(Juan et al. 2016), and CFM(Li, Do, and Liu 2019). The comparisons of their performance are reported in Table I. From the low p-values at the bottom row of the table we can see that the SFM model significantly outperforms other alternative models.

Conclusion

In this research we propose a Spatial Factorization Machine (SFM) model to address the problem of traffic status estimation on un-monitored roads. Our SFM model takes geo-graphical information such as distance matrices into the training process and optimizes the coupled matrices via commonly factorized latent factors. Experiments on real world traffic data demonstrate that the SFM model is significantly better than other existing models in traffic estimation. In future, we will extend the SFM model to other spatial estimation problem such as point of interest recommendations.

References

Bellini, P.; Bilotta, S.; Paolucci, M.; Soderi, M.; et al. 2018. Real-time traffic estimation of unmonitored roads. In *16th IEEE International Conference on Pervasive Intelligence and Computing*.

Du, Z.; Huang, L.; Wen, Y.; Xiao, C.; and Zhou, C. 2018. Route planning for vessels based on the dynamic complexity map. In *Information Fusion and Intelligent Geographic Information Systems*. Springer.

Huang, H.; Bucher, D.; Kissling, J.; Weibel, R.; and Raubal, M. 2018. Multimodal route planning with public transport and car-pooling. *IEEE Transactions on Intelligent Transportation Systems*.

Juan, Y.; Zhuang, Y.; Chin, W.-S.; and Lin, C.-J. 2016. Field-aware factorization machines for ctr prediction. In *Proceedings of the 10th ACM Conference on Recommender Systems*.

Li, L.; Do, Q.; and Liu, W. 2019. Cross-domain recommendation via coupled factorization machines. In *the Thirty-Third AAAI Conference on Artificial Intelligence (AAAI)*.

Rendle, S. 2010. Factorization machines. In *the 2010 IEEE International Conference on Data Mining*.