COLLABORATIVE LEARNING FOR DEMAND FORECASTING IN URBAN LOGISTICS

Maharshi Dhada[1], Matteo Giaretti, Duncan McFarlane[1]
[1]Institute for Manufacturing, Department of Engineering, University of Cambridge, Cambridge, CB3 0FS, UK.

Abstract

Purpose: This paper presents a probabilistic hierarchical statistical model for tackling the often encountered cold-start problem in urban logistics sector. Companies face the cold-start problems when they start their operations in a new city and therefore don't have sufficient data to accurately forecast the demand for that city. Using real-world industrial data, authors propose a hierarchical autoregression model that enables a logistics company to forecast customer orders in a new city, given the historical orders in other cities.
Design/ methodology/ approach: Using a statistical hierarchical model enables the user to systematically model the operations of an urban logistics company across multiple cities. Each city is associated with an individual autoregression model, whose parameters are sampled from a common higher level distribution that represents the general behaviour of other similar cities. Operations in a new cities are therefore enhanced using systematically evaluated prior knowledge, enabling reliable forecasts in the early time-steps. Similarities across the cities are modelled using factors such as population density, geography, cultural influences, etc.
Findings: Using a statistical hierarchical model enables the user to systematically model the operations of an urban logistics company across multiple cities. Each city is associated with an individual autoregression model, whose parameters are sampled from a common higher level distribution that represents the general behaviour of other similar cities. Operations in a new cities are therefore enhanced using systematically evaluated prior knowledge, enabling reliable forecasts in the early time-steps. Similarities across the cities are modelled using factors such as population density, geography, cultural influences, etc.
Originality/ value: This is the first application of statistical hierarchical modelling for forecasting customer demand in urban logistics sector. Such a cold-start problem addressed in this paper is critical for several urban logistics companies, and the results shown in this paper are evaluated using real-world industrial dataset which justify the applicability of the technique.

Keywords: Machine Learning, Logistics, Collaborative Learning, Statistics, Hierarchical Modelling

Introduction

Urban logistics sector aims at meeting consumer expectations summarised in Carlos Moreno's idea of a '15-minute city', where essential goods and services are accessible within 15 minutes of a resident's home in an efficient and environmentally sustainable manner (Moreno et al., 2021) The rise in the quick e-commerce segment, where the customers expect rapid and free deliveries, has been further accelerated by the Covid-19 pandemic. The lockdowns in particular restricted the public movement, meaning many products and services could only be availed online. As such, high prevalence of delivery vans, HGVs, motorcycles, or bicycles, delivering goods or food from one part to the other can be seen on the streets of major cities today (ARUP, 2023; Statista Research Department, 2023).

A recent ARUP report highlights this boom in urban logistics sector through the current decade. The report presents that 66% of millennials routinely look for one- hour delivery options, with the 'Amazon effect' leading to more consumers expecting free deliveries. The same-day or instant delivery service is expected to grow by 20-25% (ARUP, 2023). And consequently, the number of delivery vehicles on the roads of the 100 most significant global cities is expected to increase by 36% from 2019 to 2030 (World Economic Forum, 2020).

This paper discusses a particularly interesting problem faced by a technology company Glovoapp23 SA, popularly known as Glovo, which serves as a platform to bring together couriers, individuals, and organisations in urban areas. Glovo offers their customers an exclusive home delivery service, not just for food but delivering anything you could possibly need, from pharmacy products and gifts to flowers, contact lenses, and much more (Glovoapp23 SA, 2021).

A key driver of Glovo's operations is their forecasting engine that predicts the number of customer orders, enabling Glovo to ensure optimal availability of the couriers. However, forecasting customer orders is challenging for the cities where sufficient data does not exist for training the forecasting algorithms, often these cities are where Glovo operations have newly commenced. Such a cold-start problem can also be seen for events which are not strictly seasonal, presence of a new competitor in the market, or reasons that have a long lasting effect on the demand such as Covid-19.

This paper discusses and shows, with an example of a hierarchical auto-regressor model, that there exist opportunities for the cold-start problem to be solved by selectively learning from other cities where Glovo operates. The initial results show that an isolated forecasting model for a particular city with insufficient data is associated with a high variance, whereas a forecasting model common to all the cities is associated with a high bias. A hierarchical forecasting model, where the parameters of isolated forecasting models are sampled from a single overlying distribution common to the cities with similar behaviour, systematically enables cross learning via combined inference using the data obtained from the cities. More generally, this provides for an industrial strategy where the lack of local information can be alleviated via collaborative learning within the system.

The following paper is as follows: Section 2 Discusses the existing industrial applications of hierarchical modelling and collaborative prognosis technique, for example industrial collaborative prognosis. Section 3 provides a brief introduction to Glovo's forecasting problem for cities with insufficient data. A hierarchical auto-regressor model that was implemented for cross-learning at Glovo is discussed in Section 4. The results are discussed in Section 5.

Hierarchical Models for Cross-Learning

The term knowledge transfer is used in the probabilistic machine learning literature to refer to methods that learn from multiple related datasets. While the multi-task approach also assumes the predictors (i.e. tasks) are correlated over the fleet, the parameters across domains are learnt at the same time with equal importance. Hierarchical Statistical Modelling offers a multi-task framework, where a model is built with a 'hierarchy' of parameters, whereby domain-specific tasks are correlated via shared latent variables (Bull et al., 2022)

For the industries, statistical hierarchical models enable combined inference from the asset fleet data by learning a set of correlated models via shared higher level distributions. The shared higher level distributions ensure that the behaviour observed across other similar assets in the fleet is incorporated when the assets have sparse data. The parameters of the overall model are learnt using hierarchical Bayesian inference and provide robust variance reductions compared to the independent models inferred for the as- sets with sparse data. Hierarchical models therefore automatically incorporate collaborative learning across similar assets or sub-fleets, such that the assets with sparse data borrow statistical strength from those that are data-rich. Comprehensive information about statistical hierarchical modelling can be found in (Gelman et al., 1995)

One of the earliest applications use hierarchical models for inferring the Bernoulli parameters for reliability estimation of emergency diesel generators in separate nuclear power plants. They show that the hierarchical Bernoulli model was more accurate for modelling the collective "composite" and individual reliabilities of the generators, compared to the prevalent approach of analysing data from all generators as a single dataset. (Di Francesco et al., 2021) use hierarchical models to build corrosion models given the data from multiple sources. An interesting application can also be found in (Johnson et al., 2005) where hierarchical modelling was used for reliability estimation of new space crafts, which had only experienced few failures or in some cases no failures.

Glovo's Forecasting Problem

Forecasting lies at the core of Glovo's data science to optimise its operations, along with two other fundamental engines which are matching and routing. Glovo's end-to-end operations include the following steps:

1. **Forecast** the number of orders in a city for a given time. This is to open vacancy for the optimal number of couriers in the city
2. **Match** the orders with the optimal courier out of the ones available at that instance

3. **Route** the courier journey so that it can optimally pick up the parcel and deliver it to the customer

This paper focuses on the forecasting step where the problem is to forecast the number of orders, given the past orders. Glovo operates in hundreds of cities from which the past orders training data can be obtained. An example of the daily orders across a subset of cities is shown in Figure 1, where each city is indicated with a different colour. It is apparent from the figure the cyclic pattern which peaks at dinner and lunch times.

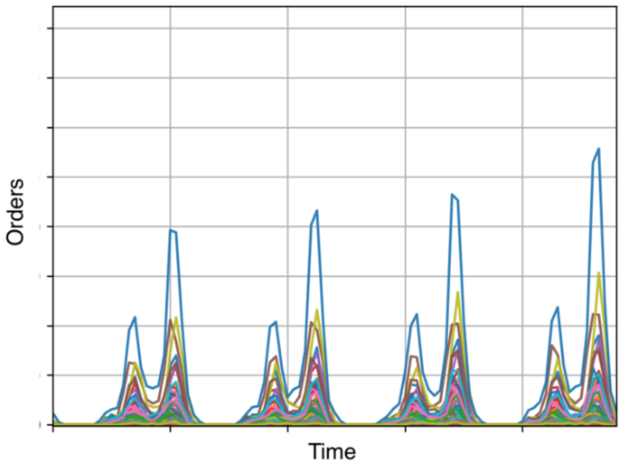

Figure 1: Example of daily orders from a subset of cities

Upon analysing the partial auto-correlation function (PACF) values, an auto-regressor with lag 2 (represented henceforth as AR(2)) is deemed suitable model for forecasting and for the purpose of this paper. The mathematical representation of this model is shown in (1), where the customer orders forecast for city i, at the current timestep t, is linearly correlated with the two previous timesteps and a noise coefficient ϵ.

$$X_{i,t} = \alpha_{i,1}X_{i,t-1} + \alpha_{i,2}X_{i,t-2} + \varepsilon \quad (1)$$

An example of this forecasting model, showing true vs. predicted customer order numbers for a city is presented in Figure 2, giving a MAPE score of 16.3% across all cities. This shows that the AR(2) model is sufficient for forecasting with acceptable accuracy, if sufficient training data is available.

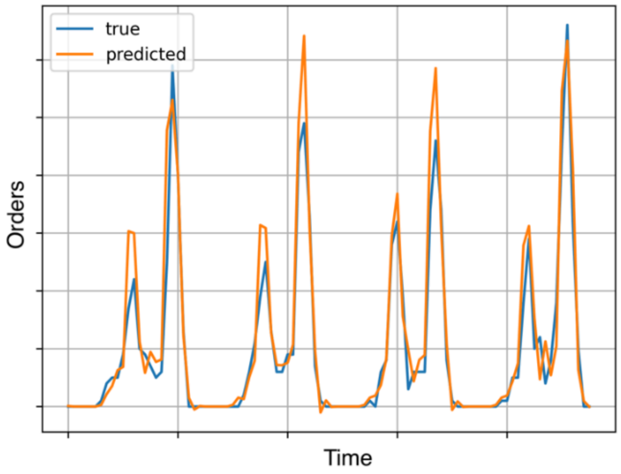

Figure 2: True vs. predicted order numbers using AR(2)

However, the auto-regressor ceases to be reliable when the training data is insufficient. A probabilistic AR(2) was used to highlight this drawback, where the model coefficients were sampled using Monte Carlo Markov Chain (MCMC) sampling technique and the probabilistic programming language Stan (Carpenter et al., 2017). Time- series of a randomly selected city was truncated to simulate insufficient training data, and gradually increased. The probabilistic estimates of the AR(2) model coefficients were recorded as distributions as the timeseries data was increased, to study the effect of training data on the predictive accuracy and confidence. The accuracy was compared evaluated AR(2) coefficient values trained using the same timeseries but with a large enough training dataset to achieve stable values, and the confidence was evaluated based on the variance (or spread) of the inferred distributions of the AR(2) coefficients.

The results of this exercise are shown in Figure 3. It is evident from this figure that when there is lack of training data, the estimates are (1) inaccurate and (2) associated with high variance. Here the thick vertical line in the plots correspond to their true values.

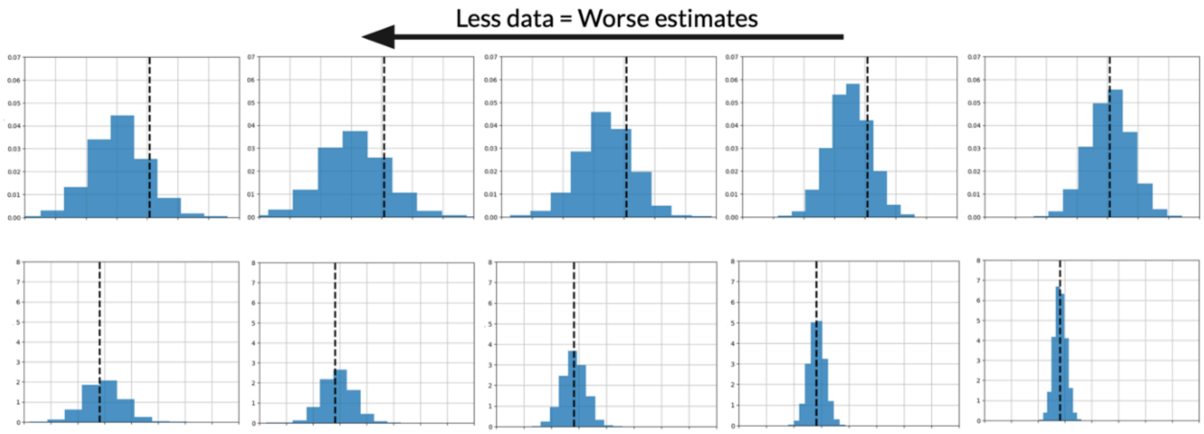

Figure 3: Low data resulting in inaccurate inference of the AR(2) coefficients, with high variance

Hierarchical Model for Glovo Forecasting

In this section, the methodology to experiment with a hierarchical AR(2) model is described. The aim of this experiment was to establish whether a hierarchical model is better suited for forecasting when a given city does not have sufficient training data.

Hierarchical Model

In the hierarchical model for forecasting, for the problem at hand, clusters of cities that show similar trends shall share higher level distributions from which the city-specific AR(2) slope coefficients are sampled. For this case only α parameters shared higher level distributions as the noise variable is treated as random, and in case of a hierarchical model it should ideally be shared across all cities similarly. Concretely, the α are sampled from higher level Gaussians, with noninformative priors having high corresponding variances. This is represented in mathematical terms in (3), where the j^{th} alpha coefficient is sampled from the corresponding Gaussian N_k shared by the cluster k comprising of cities showing a similar behaviour.

$$\alpha_{i,j} \sim N_k(\mu_j, \sigma_j) \qquad (2)$$

$$X_{i,t} \sim N\left(\alpha_{i,1}X_{i,t-1} + \alpha_{i,2}X_{i,t-2}\right) + \epsilon \qquad (3)$$

Design of the Experiment

In an ideal setting, the clusters of cities with similar trends should be identified using meta data such as the geography, weather, presence of other competitors, population density, etc. But the focus of this paper is to show that a hierarchical model improves the prediction for a city with insufficient data, in the presence of other similar cities with higher information. The clustering is not the focus, but it can nonetheless be coupled in the inference steps via algorithms such as expectation-maximisation.

The same subset of cities shown in Figure 1 are used for the experiment discussed here. There exists sufficient data for each of these cities to train an AR(2). To identify the clusters of similar cities, standard k-means algorithm was implemented using the coefficients of the AR(2) models trained for each city using the complete training dataset. The city clusters identified through this procedure are shown in Figure 4, where the city clusters are shown in various colours.

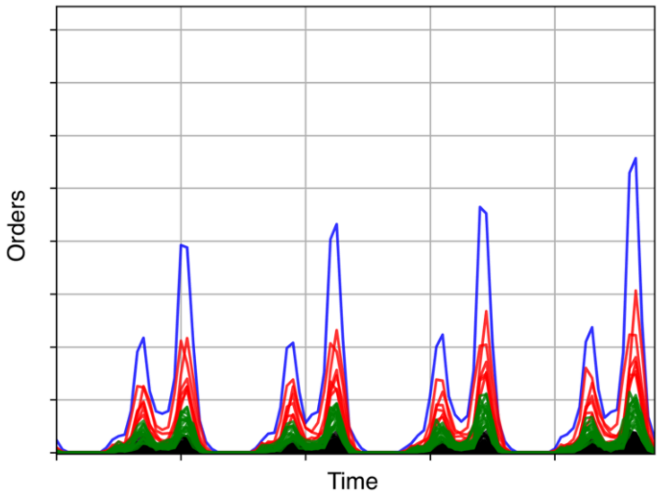

Figure 4: City clusters identified and shown in difference colours

Once the clusters are identified, the same city selected for plotting the values in Figure 3 was used to experiment with the hierarchical AR(2) model. The data was increased in the same quantities and iterations as Figure 3, with the only difference being this time the AR(2) model coefficients of this city were sampled from a distribution shared among other cities in its cluster which had more training data. The distributions obtained after inferring both the slope coefficients of the hierarchical AR(2) model are shown in Figure 5.

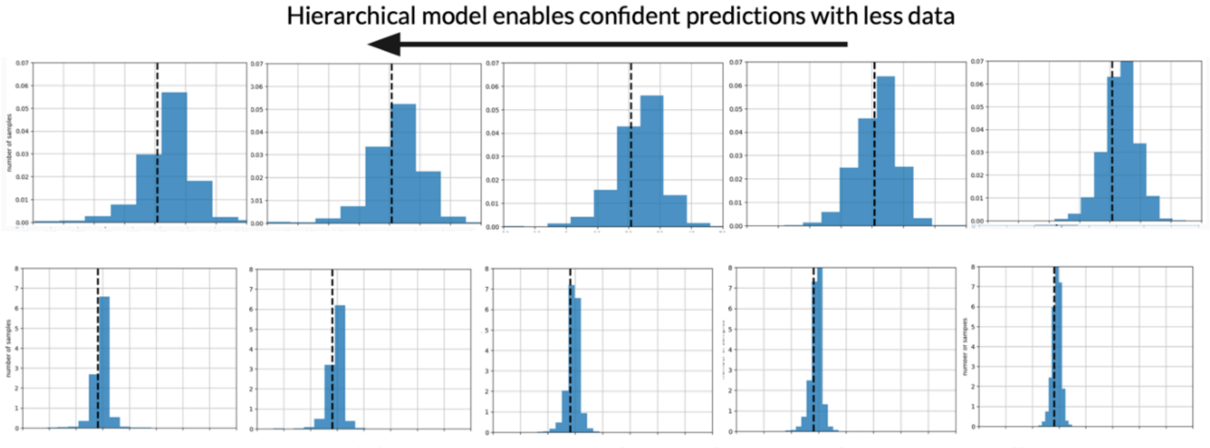

Figure 5: Hierarchical AR(2) model enables confident inferences of the slope coefficients, despite insufficient training data

Discussion and Results

The initial results presented in this paper show that, for the case of Glovo, an isolated AR(2) model for cities with less data is associated with a high variance. In theory, an AR(2) common to all cities in the Glovo operations shall be associated with a high bias, given the range of customer order values. The results in the Figure 5 show that a hierarchical AR(2) balances the above by systematically incorporating prior knowledge from other cities showing similar trends.

The future research comprises of identifying the meta data and a systematic algorithm for identifying the clusters of cities showing similar trends, and also a forecasting model that incorporates seasonal long-term variations in the consumer orders.

Acknowledgements

This research was supported by the SF Express Research Fellowship in Logistics. This research was also made possible by Glovoapp23 SA, based in Barcelona.

References

- ARUP. (2023). *Future of Urban Logistics*. Https://Www.Arup.Com/Insights/Future-of-Urban-Logistics/. https://www.arup.com/insights/future-of-urban-logistics/
- Bull, L. A., Di Francesco, D., Dhada, M., Steinert, O., Lindgren, T., Parlikad, A. K., Duncan, A. B., & Girolami, M. (2022). Hierarchical Bayesian modeling for knowledge transfer across engineering fleets via multitask learning. *Computer-Aided Civil and Infrastructure Engineering*.
- Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., & Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, *76*(1).
- Di Francesco, D., Chryssanthopoulos, M., Faber, M. H., & Bharadwaj, U. (2021). Decision-theoretic inspection planning using imperfect and incomplete data. *Data-Centric Engineering*, *2*, e18.
- Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (1995). *Bayesian data analysis*. Chapman and Hall/CRC.
- Glovoapp23 SA. (2021). *How does Glovo work? The answers to all your questions*. Https://Medium.Com/Glovo-Engineering. https://medium.com/glovo-engineering
- Johnson, V. E., Moosman, A., & Cotter, P. (2005). A hierarchical model for estimating the early reliability of complex systems. *IEEE Transactions on Reliability*, *54*(2), 224–231.
- Moreno, C., Allam, Z., Chabaud, D., Gall, C., & Pratlong, F. (2021). Introducing the “15-Minute City”: Sustainability, resilience and place identity in future post-pandemic cities. *Smart Cities*, *4*(1), 93–111.
- Statista Research Department. (2023). *Last-mile logistics worldwide - Statistics & Facts*. Https://Www.Statista.Com/Topics/4383/Last-Mile-Delivery/. https://www.statista.com/topics/4383/last-mile-delivery/
- World Economic Forum. (2020). *The Future of the Last-Mile Ecosystem*. The Future of the Last-Mile Ecosystem. The Future of the Last-Mile Ecosystem.