EPSRC Centre for Doctoral Training in Industrially Focused Mathematical Modelling

Customer mobility and congestion in supermarkets

Fabian Ying

Contents

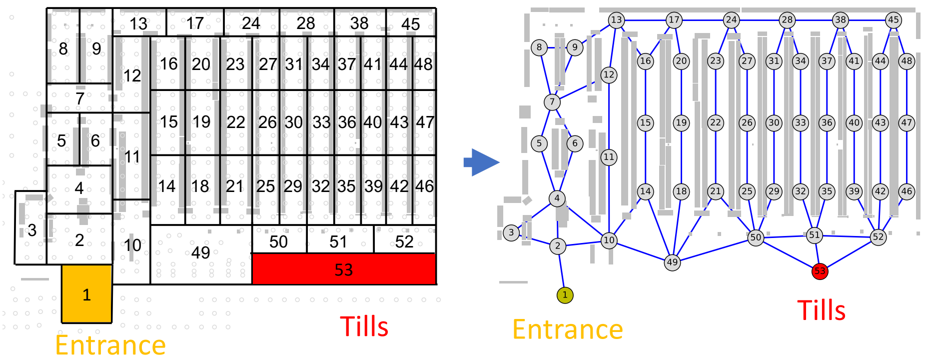


Figure 1 – Representing a store (left) as a network (right)

1 Introduction

Background

Understanding how customers move inside stores is of considerable interest to retailers, as it can help them improve store layouts with less congestion, more item sales, or other desirable features. However, the mobility of customers in a store is a complicated process that depends on many factors, including customers' item preferences, shopping lists, familiarity with a store's layout, the availability and price of items, and the level of congestion in a store. Some of these factors also vary over time (e.g., availability of items or congestion), and the flow of people or the way in which people navigate therefore varies over time. A mathematical model that explains the main features of customer mobility and that gives good estimates of the movements (even if on an aggregate level) is of great interest to retailers. Such a model can be used to improve store layouts, so that there is less congestion or greater exposure of promotional items.

Our work is motivated by the problem of understanding how customers move in a supermarket and the problem of measuring and mitigating congestion inside a store. There are three primary objectives:

- model how customers move in a supermarket,
- model congestion based on customer movements, and
- identify store layouts that minimize congestion.

Congestion negatively affects customer satisfaction and delays staff in fulfilling online orders.

Supermarket companies seek to reduce congestion inside their stores, because it has a negative effect both on customer shopping experience and on the fulfilment time for online orders. In many supermarkets, staff members go around a store (at the same time as customers shop in the store) and pick up items that were ordered online. Congestion may delay such orders and thereby incur additional costs and inconvenience customers in the stores.

Approach

Our approach is to represent a store as a network (which we call *store network*), in which each node represents a zone in a supermarket and each edge connects adjacent zones (see Figure 1).

We use state-of-the-art population-level mobility models to estimate the mobility flow of customers between zones in a supermarket. These models take the store network and the popularity of zones (measured roughly by the number of purchases made in each zone) as input and return an origin–destination (OD) matrix $\boldsymbol{T}$ with entries T_{ij} which record the mobility flow from zone i to zone j during a time period (3 months in our case). We measure the mobility flow T_{ij} by the number of times that a customer purchased an item in zone i followed by a purchase in zone j.

We then combine the mobility model with a congestion model to estimate congestion in a supermarket. Our congestion model is based on queueing networks, in which each node

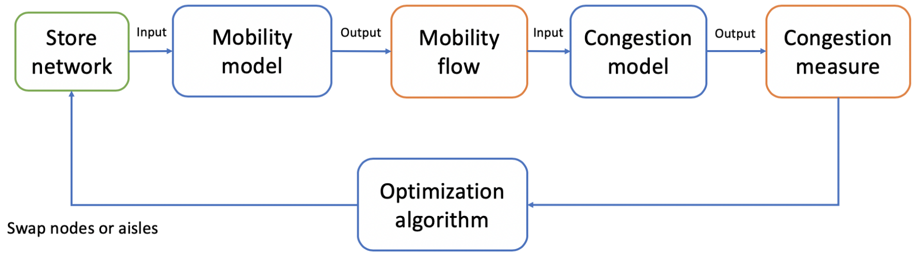


Figure 2 – Summary of our approach

(or zone) acts as a queue, at which customers queue up to be served. Such a queueing network framework allows us to estimate the mean time it takes for customers to traverse a node based on the number of other customers in the same node. We measure congestion by the mean time it takes for customers to finish their shopping journey. Finally, we use an optimization algorithm to find better store layouts with lower values of our congestion measure. We summarize our approach in Figure 2.

Glossary of terms

- Store network: Network representation of a store
- Origin–destination (OD) matrix: Matrix $\boldsymbol{T}$ with entries T_{ij} which record the mobility flow from zone i to zone j during some time period.
- Common part of commuters (CPC): Proportion of trips that are in both the empirical OD matrix and OD matrix that is estimated from a model.

2 Data

We use anonymized ordered basket data to estimate the mobility flow in supermarkets.

Our data consists of anonymized ordered customer-basket data from 17 Tesco stores over a common three month period. Each ordered customer basket is a list of item purchases, which we order by pick-up time. We use item-location data to map each ordered list of purchases to their associated zones in a supermarket. We calculate the empirical origin–destination matrix $\boldsymbol{T}^{\text{data}}$ by counting the number of times that customers bought an item in zone i followed by a purchase in j for every pair of nodes (i, j).

3 Mobility models

We use mobility models which have been used to estimate flow of commuters, goods, and vehicle traffic.

We use the following population-level mobility models to estimate the mobility flow in a supermarket.

- Gravity model: Inspired by Newton's law of gravity, the mobility flow under the gravity model is proportional to the popularities of the origin and destination nodes and is inversely proportional to the distance (to some power α) between these two nodes (see Figure 3). The gravity model has been used for many decades in various applications such as estimating flow of commuters, flow of trade goods, traffic flow, and identifying catchment area of shops, hospitals, and schools.
- Intervening–opportunities model: The main idea of the intervening–opportunities (IO) model is that the mobility flow between two nodes i and j increases with the popularities of the origin and destination nodes and decreases with the total popularities of the *intervening* nodes associated with the pair (i, j). An intervening node in this case are all nodes that are closer to i than j is. These nodes are perceived as more desirable, as it is closer to travel to the intervening nodes from i than to travel to j. Therefore, in contrast to the gravity model, the distance between the two nodes i and j only indirectly affect the mobility flow between them: The larger the distance between two nodes, the more intervening nodes are there, so the total

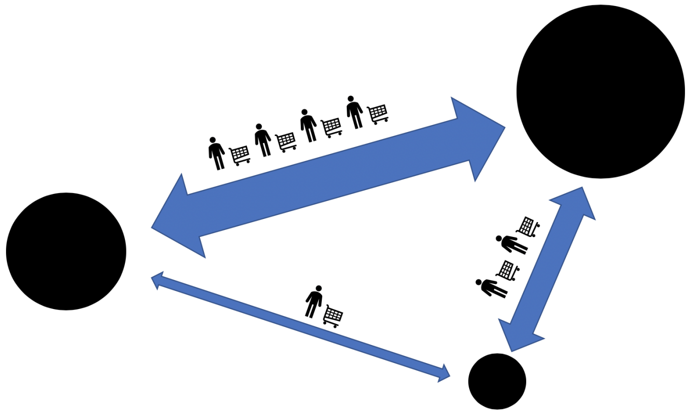


Figure 3 – Illustration of the gravity model. The larger the popularities of the nodes (black circles) and the shorter the distance between them, the larger the mobility flow between them.

popularities between the two nodes increases and there is a smaller mobility flow between them under the IO model.

- Radiation model: Inspired by wave theory, the radiation model is a parameter-free variant of the IO model.
- Extended radiation model: This model is an extension of the radiation model that includes an additional calibration parameter that has been reported to give better fits on smaller spatial scales.

In our application, the popularity of a node correspond roughly to the number of purchases made in each node.

Goodness-of-fit measures

Each of the mobility models outputs an origin–destination matrix $\boldsymbol{T}^{\text{model}}$, and the goal of each model is to output an OD matrix $\boldsymbol{T}^{\text{model}}$ that is as 'close' as possible to the empirical OD matrix $\boldsymbol{T}^{\text{data}}$. However, we require a suitable notion of 'closeness' when comparing these two matrices. For this purpose, we use the following two goodness-of-fit measures:

- Common part of commuters (CPC): The common part of commuters measures the proportion of trips that occur in both $\boldsymbol{T}^{\text{model}}$ and $\boldsymbol{T}^{\text{data}}$. In other words, it is the proportion of trips that were estimated correctly by the model.
- Error in estimated number of visits: When we consider congestion in Section 5, we estimate the number of visits to each node from an OD matrix $\boldsymbol{T}$ by assuming that customers take shortest paths between purchases. We separately calculate the estimated number of visits from the model estimate $\boldsymbol{T}^{\text{model}}$ and from the empirical OD matrix $\boldsymbol{T}^{\text{data}}$. We calculate the discrepancy in the estimated number of visits calculated in these two ways by the normalized root-mean-square error in these values.

Parameter calibration

The gravity, IO, and extended radiation model each have a single fitting parameter. We calibrate these parameters by maximizing CPC. In other words, for each model, we choose the parameter such that the model achieves the highest value of CPC.

4 Results of fitting mobility models

The gravity model gives the best fit to the empirical mobility flow. It estimates on average 69% of the flow.

We apply the four mobility models from Section 3 to 17 stores. The gravity model performs the best both in terms of the CPC score and the error in estimated number of visits (see Table 1). The CPC values of the gravity model are also comparable to the performance of this model in previous applications such as estimating commuting flow. The error in the estimated number of visits is also low for the gravity model (see also Figure 4). Therefore, the gravity model describes well the mobility flow in supermarkets. We further found that the model parameters of the gravity model does not change significantly between different stores, so one can use the model parameter from one store to estimate the mobility flow in all other stores. This observation is particularly useful in practice for estimating the mobility flow in stores for which we do not have empirical data on the mobility flow, as we would normally require data on the empirical mobility flow to calibrate the model parameters.

Table 1 – Mean CPC scores and error in estimated number of visits for the four mobility models. We highlight the best value in each column in bold.

Model	Mean CPC (higher better)	Mean error in estimated number of visits (lower better)
Gravity	**0.686**	**0.045**
Ext. radiation	0.672	0.054
IO	0.655	0.047
Radiation	0.513	0.116

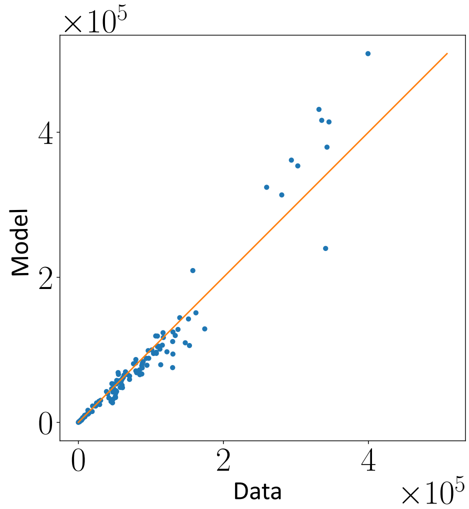


Figure 4 – Comparison between the estimated number of visits to each node calculated from the OD matrix from the gravity model (Model) and from the empirical OD matrix (Data). The orange line is the identity. The gravity model agrees well with the data in terms of the estimated number of visits to each node.

So far, we have shown that the gravity model describes well the mobility flow in supermarkets and that its parameter value does not change significantly between different stores (which each have different store layouts). Therefore, we expect the gravity model to work well in estimating the change in mobility flow when changing the store layout. Given a suitable model for how customers walk between purchases, one can estimate shelf-level foot traffic and congestion from the mobility flow. (Note that the mobility flow does not make any assumptions on how customers walk between purchases.) Our models therefore allow supermarket companies to test out different store layouts and estimate the mobility flow (and therefore the foot traffic and congestion) for each of them, so they can select a layout with more desirable properties than the existing layout. We demonstrate one application (estimating and mitigating congestion) in the next section.

5 Integrating mobility model with congestion model

We integrate our mobility model with a congestion model based on queueing networks to estimate congestion in supermarkets.

To estimate congestion from a mobility flow (which we estimate using the gravity model), we assume that customers traverse shortest paths between purchases and that every node acts as a queue with a single server (see Figure 5). Each customer queues up at each node to be served. The server at each node serves at a rate μ (the service rate). In practice, the service rate μ can be calibrated using empirical data on the mean traversal time of each node. However, as we do not possess this data, we choose a value of μ that is deemed plausible (in particular, which gives plausible values for the mean traversal time). We estimate congestion by the total mean queue size Q, which is the mean number of customers in the store. Minimizing Q is also equivalent to minimizing the mean journey time of customers.

Figure 5 – Diagram of a queue. We use queues to model congestion at each node.

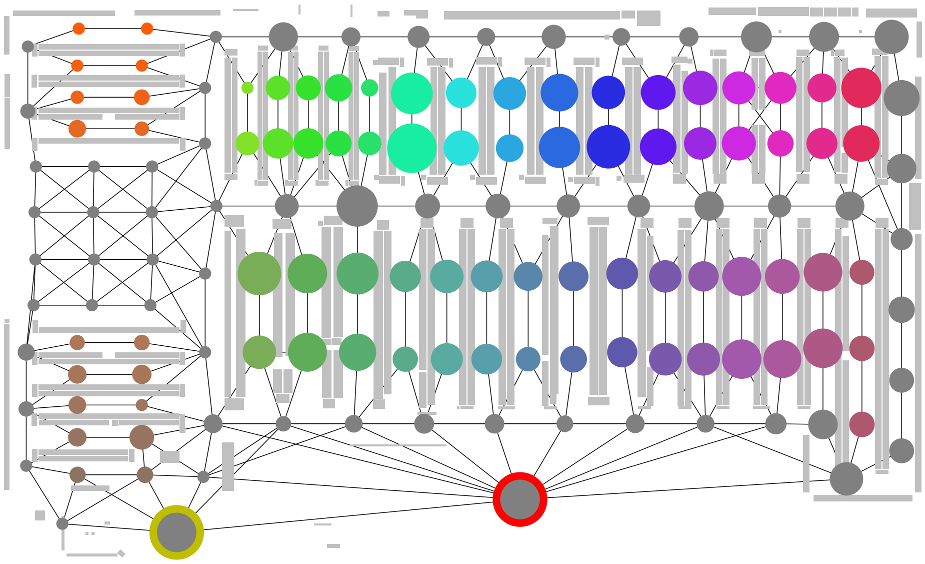
We use an optimization algorithm (called simulated annealing algorithm) that swaps aisles with one another to find store layouts that reduce our congestion measure Q. Our algorithm is able to find store layouts with significantly lower values of Q (around 25% lower) than the original store layout. In the store layout with the smallest found value of Q, popular nodes were moved from the centre of the store towards the perimeter of the store (see Figure 6).

6 Summary and future directions

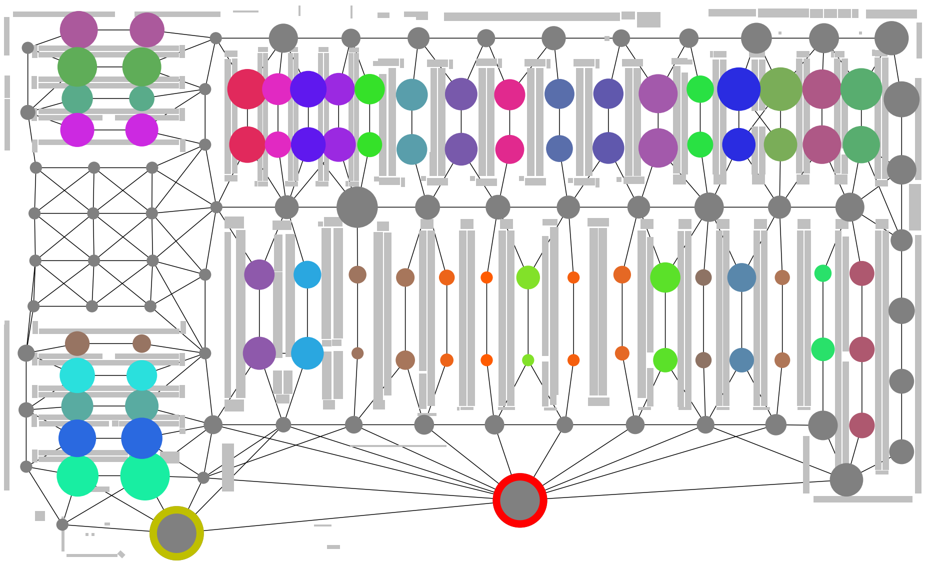
In summary, we considered the problem of modelling and analyzing customer mobility and congestion in supermarkets. We were motivated by the problem of modelling how customers move in a supermarket and finding an optimal supermarket layout that minimizes congestion. In our approach, we represented a store as a spatial network (which we called a "store network") in which the nodes are zones of the store and edges connect adjacent zones. We estimated the empirical mobility flow in 17 supermarkets from anonymized and ordered customer-basket data. We fit the mobility models to this data and found that the gravity model successfully estimate 65–70% of the flow inside supermarkets. Being able to estimate mobility flow in a supermarket has potential applications for estimating shelf-level foot traffic and congestion in supermarkets, as well as finding store layouts with redirected foot traffic or less congestion. We demonstrated the latter application by combining the gravity model with a congestion model based on queueing networks to estimate congestion. We applied an optimization tool to it to identify store layouts with lower values of congestion.

Our work has potential applications in designing more efficient and less congested stores.

Our work is not without limitations. These mainly concern on the simplifying assumptions that we made. For example, we assume that customers traverse shortest paths between purchases. However, in reality, customers traverse paths that deviate from a shortest path. Furthermore, our congestion model based on queueing networks has not been verified empirically, so all estimates of congestion may not be accurate. However, our approach does not rely on the specific choices of model for how customers traverse between purchases and congestion models. In future work, one can replace them with more



(a) Original store layout



(b) Optimized store layout
when minimizing the total mean queue size Q

Figure 6 – Location of popular nodes before and after optimization. Nodes of the same colour belong to the same aisle. Gray nodes do not belong to any aisles. The node size is proportional to its popularity (roughly the number of items sold at the node). We circle the entrance and till nodes in yellow and red, respectively.

accurate models to obtain better estimates of congestion or other quantities of interest. Our approach is therefore a first step towards modelling customer mobility and congestion in supermarkets.

For more details on this work, see [1, 2].

Comment from Alisdair Wallis, Data Science Manager at Tesco

Dr. Alisdair Wallis, Data Science Manager at Tesco, said: "..."

References

[1] Fabian Ying, Alisdair OG Wallis, Mariano Beguerisse-Díaz, Mason A Porter, and Sam D Howison. Customer mobility and congestion in supermarkets. *arXiv preprint arXiv:1905.13098*, 2019.

[2] Fabian Ying. *Customer mobility and congestion in supermarkets*. D.Phil. thesis, University of Oxford, 2019.