Efficient parcel delivery by predicting customers' locations

Reference:
Praet Stiene, Martens David.- Efficient parcel delivery by predicting customers' locations
Decision sciences - ISSN 0011-7315 - 51:5(2020), p. 1202-1231
Full text (Publisher's DOI): https://doi.org/10.1111/DECI.12376
To cite this reference: https://hdl.handle.net/10067/1592010151162165141

Efficient parcel delivery by predicting customers' locations

Abstract

An important aspect of the growing e-commerce sector involves the delivery of tangible goods to the end customer, the so-called last mile. This final stage of the logistics chain remains highly inefficient due to the problem of failed deliveries. To address this problem, delivery service providers can apply data science to determine the optimal, customer-centered location and time window for handover. In this paper, we present a three-step approach for location prediction, based on mobile location data, in order to support delivery planning. The first step is identifying a user's locations of interest through density-based clustering. Next, the semantics (home or work) of the user's locations of interest are discovered, based on temporal assumptions. Finally, we predict future locations with a decision tree model that is trained on each user's historical location data. Though the problem of location prediction is not new, this work is the first to apply it to the field of parcel delivery with its corresponding implications. Moreover, we provide a novel and detailed evaluation on real-world data from a parcel delivery service. The promising results indicate that our approach has the potential to help delivery service providers to gain insights into their customers' optimal delivery time and location in order to support delivery planning. Eventually this will decrease last-mile delivery costs and boost customer satisfaction.

INTRODUCTION

An increasing number of customers order a broad range of products via online delivery rather than going to a physical store. Worldwide retail e-commerce revenues rise annually and are projected to grow from 1.92 trillion US dollars in 2016 to 3.4 trillion US dollars in 2019 (Statista, 2016). In Belgium, around 67% of the population made an online purchase in 2018 and home delivery was mentioned as one of the main drivers to purchase online (Comeos, 2018). However, physical delivery of ordered products to the end consumer, also known as the *last mile*, remains a challenge and is currently regarded as one of the least efficient and most polluting parts of the logistics chain (Gevaers, Van de Voorde, & Vanelslander, 2011). One of the main reasons for this is the large portion of *failed deliveries*, when a parcel could not be received, for example because the customer was not at home (Arnold, Cardenas, Sörensen, & Dewulf, 2016). This causes inconvenience to both the delivery service provider, in terms of extra costs, miles and emissions (Gevaers et al., 2011), and the customer. A possible solution is to deliver parcels to a physical depot where the customer can collect his or her order. However, the burden of the last mile is now placed upon the customer and will reduce overall customer satisfaction. Another solution, which sounds trivial, is to ask the customer beforehand where and when they would prefer to receive the parcel. This turns out to be a double headache: one for the customer as, once such data is entered, they are bound to the appointment; the other headache for the delivery service provider, as they need to organize for more couriers to cover likely heavy peak moments. Reversely, instead of asking the customer, by determining the possible delivery times and locations, the delivery service provider can subsequently opt for the most cost-efficient solution. The optimal delivery time and location are suggested to the customer, who only needs to confirm and enjoy a hassle-free delivery.

To determine the optimal location and time window for future handover, data science can be applied. In the literature, this is done by amongst others (Pan et al., 2017) who estimated the absence probability of a customer by mining electricity consumption data, in order to improve the success rate of home deliveries in the e-grocery sector or by Van Duin, De Goffau, Wiegmans, Tavasszy, and Saes (2016) who applied address intelligence to improve home delivery efficiency for DHL. By predicting where and when a customer wants to receive a parcel, a delivery route can be planned that is both minimizing transportation costs while maximizing customer satisfaction.

The goal of this research is to predict a weekly schedule of a customer's locations, based on mobile location data, in order to support delivery planning. Since human mobility is believed to be characterized by a certain degree of regularity, such as daily and weekly routines, future locations are predictable to some extent (Gonzalez, Hidalgo, & Barabasi, 2008). A three-step approach is followed to predict users' locations from historical Global Positioning System (GPS) data. The first step is to identify personally meaningful locations a user spends a lot of time. Only those GPS coordinates where a user stayed for more than 5 minutes are selected and clustered into locations using a density-based clustering algorithm. Thereafter, clusters are ranked based on the total time spent at the location, to obtain the user's locations of interest. The next step is to discover the semantics (home or work) of the user's locations of interest. Home prediction is based on the idea that a user is most likely found at home during the night, from 0h to 5h. The work location on the other hand, is where a user is most likely found during working days between 9h and 17h. In the final step, we introduce two methods to predict the probability a user will be at a certain location given a day of the week and hour of the day.

The data that was used for this study was provided by Parcify, a Belgian last-mile delivery company currently active throughout Flanders. Parcify envisions to alleviate the aforementioned last-mile delivery issues by delivering parcels whenever and wherever the customer wants. Online purchases are delivered to a Parcify hub and Parcify distributes the purchases by bike to the customer's location of choice (as indicated by the customer via a mobile application) (Parcify, 2016). This way, Parcify reduces costs of failed deliveries while optimizing customer satisfaction. The mobile application is used to gather GPS location data of the users over time, of course with a clear opt-in and privacy statement. This research provides Parcify with a deeper insight in their customers' optimal delivery time and location, which can be used to support both operational (eg. routing) and strategic (eg. parcel hub location) decisions.

The main contributions of this paper are: (i) this paper introduces a novel solution to simultaneously reduce failed deliveries and transportation costs while optimizing customer satisfaction, (ii) we offer a practicable three-step approach, where the choice of individual techniques and parameters can be tailored to the company's specific needs, together with guidelines for evaluation and implementation, (iii) we provide an in-depth evaluation on real-world location and parcel delivery data, which shows to what extent customer's delivery time and locations are predictable.

LITERATURE OVERVIEW

Thanks to the availability of various techniques to collect accurate geolocation data from mobile devices, human mobility is studied at large scale in recent years. Smartphones are able to record quality location data in a continuous manner, which offers new opportunities to discover valuable information from this type of data. In this research we will focus on three application problems in particular: discovering significant locations, semantic labeling of these locations and future location predictions. A literature review shows that these applications are already used in various domains (Table 1). Context-aware mobile devices can make more intelligent decisions based on the user's (future) location. For example, a cellphone could automatically change its settings when entering a certain location, like switching to silent mode when entering a movie theater. Other examples are navigational services and location-based advertising. One step further, location predictions could benefit recommendations and reminders, e.g. a user can be reminded to stop by a supermarket when it is predicted that the user will head home. Other applications can be found in urban planning, where insights in home and work areas and the evolution thereof are useful (Dash et al., 2014), energy efficiency and demand, disease spreading, etc. However, we will apply these methods to a new field, namely parcel delivery. Next, we will look at prior research in more detail.

[Table 1 about here.]

Discovering locations of interest

Human mobility is defined as a sequence of location visits, but only few locations are relevant or meaningful in an individual's daily life. When using GPS data, various clustering algorithms can be applied to discover spatio-temporal patterns in raw location data. These patterns reveal personally meaningful locations.

A popular clustering technique is **k-means clustering** Ashbrook and Starner (2003). The technique iteratively assigns points to the nearest cluster center and recalculates the centers. The main drawbacks of this method for location discovery are (i) the number of clusters needs to be defined beforehand, which is difficult when one does not know how many locations should be discovered and (ii) sensitivity to noise because all points will be included in the final clustering results (Zhou, Frankowski, Ludford, Shekhar, & Terveen, 2007).

Time-based clustering extracts significant locations based on time and location (Kang, Welbourne, Stewart, & Borriello, 2004). When the distance between a new coordinate and the cluster center exceeds a certain threshold d, a new cluster is formed. This is a very simple approach that does not take into account the frequency of visiting the same location. Therefore, this approach is often combined with a second density- or grid-based clustering step (Montoliu & Gatica-Perez, 2010).

Density-based clustering is based on the density of points and points that are not included in any clusters by the end of the process are considered to be noise. Clusters are defined by two parameters: the radius of the cluster (Eps) and the minimum number of points within that cluster ($MinPts$) (Ye, Zheng, Chen, Feng, & Xie, 2009). Starting from an arbitrary point, neighboring points are assigned to a cluster based on these two parameters (Ester, Kriegel, Sander, Xu, et al., 1996). This way, density-based algorithms deliver some advantages (Zhou et al., 2007): (i) clusters of arbitrary shapes can be discovered, (ii) noise or outliers are not taken into account in the final clustering results, and (iii) parameters and results are relatively robust. Because of these advantages a density-based clustering method is used in this work.

The **grid-based clustering** algorithm divides the space with a uniform grid and then merges grid cells in order to obtain clusters (Do & Gatica-Perez, 2014). The boundaries of the clusters are either horizontal or vertical, so that no clusters of arbitrary shapes can be discovered. This may decrease the quality and accuracy of clusters, despite a fast processing time (Wang & Hamilton, 2009).

Automatic location labeling

With the previous methods, a limited number of a user's most important locations can be selected out of the numerous locations an individual visits. However, these locations are still represented as GPS coordinates and have no semantic meaning. To assign semantic meaning to the locations, a possible approach is to train a multiclass classification model on the descriptive features of each location, as was done by Do and Gatica-Perez (2014) and C. Huang, Jia-ching Ying, and Tseng (2012). The descriptive features can consist of temporal information (e.g. visiting time), spatial properties (e.g. land use information) (Q. Huang & Wong, 2016), the user's behavior (e.g. the number of calls or texts) and the environment (e.g. the number of observed bluetooth devices) (C. Huang et al., 2012). However, to train a classification model using all these features, one needs labeled locations. Moreover, we do not have access to user's behavior or environment information. Therefore, we will apply a simpler approach based on temporal assumptions only. For example, Dash et al. (2014) predict home locations based on the concept of inactivity, which is defined as "no activity for more than a *threshold time*" (Dash et al., 2014, p. 2), while work location prediction is based on the idea that a user goes to work on weekdays regularly but rarely on weekends.

Location prediction

In location prediction literature, the problem of *next* location prediction given the current location is predominantly discussed. However, for our application we want to be able to predict a user's locations further in the future and create a daily or even weekly overview of a user's most likely locations. Therefore, this research focuses on *future* location prediction rather than next location predictions. Below, we provide a short overview of some of the most common techniques in literature for next and future location prediction.

Next location prediction

To predict someone's next location, existing algorithms analyze the historical sequence of locations and predict the next location in the sequence, based on the current location. An often-used technique is a Markov model (Ashbrook & Starner, 2003) that is created for each location, with the probability of transitions to every other location. In a first order Markov model, the probability of transition from location A to B is calculated as the relative frequency of trips from location A to B compared to trips from A to other destinations. When using higher (nth)-order Markov models, the probability of the next state is dependent on the current state and the previous $n - 1$ states, which can significantly increase predictive power compared to first-order models (Ashbrook & Starner, 2003). In literature, numerous variations on the Markov model are used for next location prediction, such as Mobility Markov Chain (MMC) (Gambs, Killijian, & del Prado Cortez, 2012), Hidden Markov Model (HMM) (Mathew, Raposo, & Martins, 2012) and Mixed Markov Model (MMM) (Asahara, Maruyama, Sato, & Seto, 2011).

Another group of approaches is based on sequential rules or frequent pattern mining. Ryan and Brown (2012) investigate how association rule mining, an unsupervised technique to find patterns in large data sets, can be used for location prediction. They make use of the Apriori algorithm because of its simplicity and adaptability. The Apriori algorithm finds all sets of locations that occur frequently (based on a threshold). Next, association rules with a minimum confidence are generated from these sets. Also other authors prove the applicability of pattern mining techniques for next location prediction from mobile location data, such as Morzy (2007) and Monreale, Pinelli, Trasarti, and Giannotti (2009). Other methods for next location prediction include various Bayesian network (Petzold, Pietzowski, Bagci, Trumler, & Ungerer, 2005; Lee, Lee, & Cho, 2010), neural network (Vintan, Gellert, Petzold, & Ungerer, 2004) and state predictor methods (Petzold, Bagci, Trumler, & Ungerer, 2003).

Future location prediction

The previously described algorithms predict the next location in a sequence of locations and could be extended to predict the next *n* locations. The problem with this approach however, is that when the data is sparse, the sequence will have a lot of missing elements, which will deteriorate results (Ingrid, 2011). Moreover, models that work well for short-term predictions, tend to return poor results when they are used to predict further into the future (Sadilek & Krumm, 2012). An alternative way to represent the data is in the form of (time, location) pairs. The location-predictions problem is now formulated as follows: given historical time and location data, predict the location at a given time in the future. This problem is a much less discussed issue in location prediction literature (Ingrid, 2011).

Ingrid (2011) applies a variant of a Markov model, Prediction-by-Partial Match algorithm (PPM), to timestamped location data to predict location for a given time in the future. The probability for location y occurring at timestamp x is now:

$$p(y|x) = \frac{\text{\# occurrences of location y at time x}}{\text{\# occurrences of any location at time x}} \tag{1}$$

When no occurrences can be returned for a given timestamp, the model falls back to a zeroth order Markov model and returns the most likely locations, regardless of input time (i.e. the locations of interest in general).

Sadilek and Krumm (2012) developed a method to predict the most likely location a user will be at any given time in the far future, up to even months or years ahead. Principal Component Analysis (PCA) is used to extract meaningful patterns from location data and their associations with contextual features (such as day of week and hour of the day) are learned to predict future locations. Evaluation on a massive dataset shows that their model predicts the correct location with high accuracy, even years into the future. Another approach to the problem is to predict, for a given location, *when* a user will be present at this location. Ingrid (2011) developed an unsupervised model based on Market Basket Analysis to predict arrival times at a certain location. All the data are represented as a sequence of (arrival time, location) pairs. Given a location, the model searches the historical data for matches. Finally, Krumm and Brush (2011) compute a probabilistic home attendance schedule based on observed GPS data. The probability of being away from home is calculated as a function of time of the day and day of the week and represented as a linear matrix. Their probabilistic schedule proved more accurate than the participants' own predictions of their weekly home/away schedules.

DATASET DESCRIPTION

The dataset used for this research was provided by Parcify[1]. Users that want to make use of Parcify's service need to download an application on their smartphone that facilitates delivery and pick-up of parcels. By accepting the privacy policy, the user gives Parcify consent to collect location data through their mobile device. To protect customer privacy, personally identifiable information, such as name, is removed from the dataset and users are represented by a unique user ID. Data collection through the app started from April 2016 onwards, and the number of active users in the database has increased over time (Figure 1a). With a rising number of active users, the number of datapoints increased accordingly (Figure 1b). The downfall in the number of datapoints is due to a gradual switch of data collection system when users downloaded the new app starting from January 2017. In the new system, when a user stays in the same location (e.g. a building), only one datapoint is stored with a start and end time. In the previous system, several datapoints would be stored if the user's mobile device detects movement (e.g. walking around in a building).

[1] www.parcify.com

The position history dataset used for this research covers a total time span of 14 months (April 2016 - May 2017) and a total of 1767 users. Despite this extensive dataset, the datapoints per user are sometimes limited. For new users, only a short time period of data has been collected and even for older users, large data gaps reduce the amount of datapoints per user. Gaps in the data arise when a user switches off GPS tracking for a period of time. This data sparseness is an important challenge for our approach.

Per user, each datapoint is represented by latitude and longitude coordinates for the location, together with a start and end time. From this data the time spent at the location can be calculated as the difference between end and start time. GPS inaccuracy is also provided by the mobile device and indicates the error margin of the latitude and longitude coordinates, specified in meters. An overview of the features of the dataset can be found in Table 2. Data points with missing values are omitted from the final dataset. Also, observations with a GPS inaccuracy of more than 100 meters are left out. The threshold of 100 meters was chosen to be able to include enough data while still maintaining sufficient accuracy for the intended application.

[Figure 1 about here.]

[Table 2 about here.]

METHODOLOGY

After understanding and preprocessing the data, the first step is to extract locations of interest from the raw location data. In our case, locations of interest are locations where the user spends a large part of his time, as these are potential parcel delivery addresses. Subsequently, locations are labeled with a semantic meaning. This work focuses on discovering home and work locations, as these are intuitively the most important parcel delivery locations. Finally, we investigate different methods to predict where a user will be at a given time in the future. In Figure 2 an overview of the process can be found.

[Figure 2 about here.]

Discovering locations of interest

To discover locations of interest, we will follow the approach of Montoliu and Gatica-Perez (2010). From all the initial *location points* (i.e. the raw GPS location data), only those points where a user stayed for a minimum stopping time S are selected to identify *stay points*. These stay points are then clustered into *stay regions*, using density-based spatial clustering of applications with noise (DBSCAN), with radius Eps and minimum cluster size $MinPts$. An example of this process is shown for one user in Figure 3.

The goal is to define the values for the parameters S, Eps and $MinPts$ that offer the best clustering results. Since it is difficult to evaluate clustering results by themselves, we will evaluate them based on how well they perform as input for the final task: correctly discovering home and work locations. This is measured with home and work recall and precision (which will be further described in the Evaluation Section). *Home recall* is the proportion of home locations that are discovered and *home precision* is the proportion of discovered home locations that are actual home locations. In a similar way *work recall* and *work precision* are calculated. These measures are summarized into one metric: the average F1 score. The F1 score is the weighted average of precision and recall (Bramer, 2007):

$$F1\ score = \frac{2 \times recall \times precision}{recall + precision} \tag{2}$$

We select 80% of the users with known home and work addresses to optimize and select the parameters and will report results on the other 20 % in the Evaluation Section. All combinations for $S = [0, 5, 10, 20, 30]$, $Eps = [1, 10, 30, 40, 50, 100, 200]$ and $MinPts = [1, 2, 3, 4, 5, 10]$ are evaluated to find the optimal parameters. The combination of parameters that leads to the highest average F1 score consists of a stopping time S of 5 minutes, excluding locations points generated during transportation. The optimal radius Eps is 50 meters, a smaller radius could segment one actual location into multiple locations whereas a larger radius could group multiple seperate locations together. The minimum cluster size $MinPts$ is set to 1. Hence, every data point is assigned to a cluster or forms a cluster on its own, this avoids loss of information for users with few data points.

After clustering, each cluster of stay points is represented by the geometric median. This is the point minimizing the sum of distances to all the points in the cluster and is often used as an estimator of location (Eftelioglu, 2015). Moreover, it showed to provide better results in terms of distance error to the actual location in our experiments than the centroid. Using a geocoding API, the latitude and longitude coordinates of the cluster representations are converted into addresses. Finally, for each cluster, the total number of visits is counted as well as the total time spent at the location. Based on the total time spent, the clusters are ranked from high to low importance to the customer. It is expected that locations where the user spends a lot of time are potential candidates for delivery addresses and that a user is willing to receive a parcel there.

The question now becomes "how many locations do we need to consider as *locations of interest*?" Intuitively, if we consider too little locations, we might miss out on important delivery locations for the customer (low recall), whereas too many locations might become unpractical and the lower ranked locations will become unimportant to the customer (low precision). To answer this question we experimentally compare the actual delivery addresses of 305 users to our results if we consider the top 1 to 8 locations. Again we want to maximize the F1 score, this time of delivery address recall and precision. When considering the top 4 locations the F1 score is the highest.

[Figure 3 about here.]

Labeling home and work locations

Home location prediction is based on the idea that home is where a user resides at night, between 0h and 5h. In this case study, the inactivity-based method of (Dash et al., 2014) is deliberately not followed since location tracking is sometimes activated in our dataset even when a user is inactive. Although this method would take into account shift workers, false activity would possibly deteriorate the results. Secondly, each location is assigned a score that reflects the probability of being a home location. To correct for a small number of observations, the score is calculated using a smoothed version of the frequency-based estimate, i.e. the Laplace correction:

$$p = \frac{n+1}{n+m+2} \tag{3}$$

In this equation n reflects the number of nights a user was at a certain location and m is the number of nights the user was seen at other locations. The location with the highest score is presumed to be the home location of the user. Locations with a score that is only 0.1 lower than the score of the home location are also presumed to be a home location. Again, this value is empirically optimized for the F1 score of home recall and precision within a range of 0 to 1.

The work location, on the other hand, is where a user resides during the week between 9h and 17h on average. Per location, the number of weekdays a user was there between working hours is counted and for each location a score is calculated using the above equation. The drawback of this method is that it possibly will not detect home and work locations correctly for people with irregular working hours. Again, the locations with the highest score is presumed to be the work locations, as well as locations with a score that is only 0.1 lower.

Future location prediction

For each user, the home and work address are selected as candidate delivery addresses and a third location, 'other', represents all visits to locations that fall out of these two locations. The goal is to predict for every hour and day of the week, where a user will most likely be in order to support delivery planning. A first, baseline method, is based on historical frequencies. For every hour/day slot of the week the number of visits to each location is counted. Again a score is assigned to each location by using Equation 3 with n the number of times a user was at a certain location, given hour and day of the week and m the number of times the user was seen at other locations for the given hour and day of the week. This score reflects the probability that a user will be present at a certain location. When no occurrences are available for a given time slot, the model predicts the user's most frequent place together with a probability score of 50%. This model will be referred to as the *historical counts model* further on. For the moment one-hour time slots are used, but this can easily be adapted to shorter or longer time slots.

Another model is a multiclass classification model that predicts at which of the three locations (home, work or other) a user will most likely be, given the time/day slot. In this work a decision tree model is applied because it showed a fast and accurate technique compared to the other multiclass classifiers in our benchmark on this dataset (Random Forest, Nearest Neighbor, Logistic Regression, Naive Bayes and Multilayer Perceptron). However, more advanced techniques can be applied here, tailored to the dataset and to the companies specific needs.

Per user, a decision tree model is trained on historical location data that is labeled with the 3 locations. An extra feature is introduced that indicates whether the day is a business day (1) or not (0), because we expect a user's location behavior is more similar on business days. The decision tree is pruned per user on a seperate validation set of 7 days. Different values for the minimum number of samples required at a leaf node were examined and per user the value that achieved the highest accuracy on the validation set was selected.

Evaluation

The final step in our process is the evaluation of the previous steps. When putting the approach into practice, this final step is essential in order to detect the strengths and weaknesses of the method and take appropriate measures for improvement. The next sections discuss how we have evaluated the approach both for location discovery and semantics as for the location predictions.

Location discovery and semantics

The results of the location discovery and semantic labeling are evaluated against three criteria, based on the work of (Zhou et al., 2007):

- **Accuracy**. To evaluate home and work location discovery and labeling, known home and work locations, inserted by users through the app, were used. As explained in the Methodology Section, we only consider 40% of the known locations, since 60% was used to select the optimal parameters. This results in 46 home and 16 work locations. *Home recall* is the proportion of home locations that are discovered and *home precision* is the proportion of discovered home locations that are actual home locations. In a similar way *work recall* and *work precision* are calculated.

Next to that, 5 test users who agreed to reveal information about all their locations, were asked to make a list of the locations where they spend the most time. *Recall* is the proportion of items on the list that are discovered by our algorithm. *Precision* is the proportion of discovered locations that appear in the user's list. If the discovered address lies within a radius of 100 meters from the actual address, this is counted as a match. The radius of 100 meters is chosen because this error margin is workable to start delivery planning (which is the goal of this work). However, the exact delivery address is always checked with the client before actual delivery. Nonetheless, the average error distance is only 4.3 meters for users with more than 100datapoints. Only for new or inactive users with little data the error distance will be up to 100 meters. To measure the accuracy of the home and work location labeling, the 5 test users were also asked to label their home and work locations.

- **Usefulness**. Locations that are useful for our setting are locations where the user is willing to receive a parcel. For example, a user might spend a lot of time at the gym but might not want to receive a parcel here. To evaluate usefulness, the discovered locations were compared to the parcel delivery data of 305 users. This dataset contains all the addresses where a user has received a parcel. The proportion of discovered locations that appear in the parcel delivery data is an indication for usefulness. Furthermore, the 5 test users were asked to state whether they are willing to receive a parcel or not for every location on their list. *Useful recall* is calculated as the proportion of useful locations that were discovered. *Useful precision* is the proportion of locations discovered that are useful. Again, if the discovered address lies within a radius of 100 meters from the actual address, this is counted as a match.

- **Timeliness**. Finally we want to gain insight in how much data (how many days) we need to discover and label locations. Therefore, *home recall* and *home precision* for users with a rising amount of days with data were compared, as a test to find the optimal number of days with data to use for the algorithm. We focus on the optimal amount of *days with data*, to account for the data gaps, i.e. multiple consecutive days without datapoints for a user. Additionally, we compared results for our 5 test users when using 5 months of data, 2 months of data and 2 weeks of data.

Future location prediction

To evaluate our two location prediction methods, per user the labeled data was split into a train and a test set. The test set consists of the final 7 days of the data and the models were trained on all the previous data. For the decision tree model, the final 7 days of the train set were used for validation. Only users with a minimum of 21 days of data were selected to guarantee at least 7 test days, 7 validation days and 7 train days, which resulted in a total of 398 users. The predictions on the test data were compared to the true labels to evaluate model performance. The performance metrics that were looked into are *average accuracy* (averaged over all users) and the *confusion matrix*. Finally, we looked into the optimal number of days to include into the training of a model again and compared the accuracy for users with a rising amount of days.

RESULTS AND DISCUSSION

Our approach as described in the Methodology was applied to the dataset. Below, we will discuss the results in two parts: first location discovery and semantics and secondly location predictions.

Location discovery and semantics

As proposed in the Evaluation Section the results of our location discovery and semantics algorithm will be discussed on the basis of three evaluation criteria: accuracy, usefulness and timeliness.

[Table 3 about here.]

Accuracy

A test set to evaluate home and work location detection was obtained through the Parcify app, where users could insert these addresses. Most of these addresses are concentrated in the cities of Antwerp, Brussels and Ghent, where Parcify is currently active. The metrics for evaluation show that the majority of home and work locations can be correctly detected (Table 3a). However, the precision of the predictions is lower, especially for work locations. A low precision indicates that often locations are labeled as work locations while they are not. This can partially be explained by the fact that a user can have several work addresses but only mentioned one work location (e.g. the head office). Next to that, users can spend time at different (non-work) locations during working hours. Because the user was there during working hours, these locations are classified as work locations, even though the user did not indicate them as such. Yet, these locations can still be valuable as potential delivery addresses. Furthermore, to improve results, a minimum amount of data is required in order to be able to predict home and work address accurately. This motivates further testing to find the optimal amount of data, which we will further discuss under *timeliness*. Another issue with home and work detection is when a user moves to a new home or work location. When including all the historical data into the algorithm, the old locations will be misclassified as home or work. This problem could be tackled by restricting the amount of days to include or reducing the weights put on older data.

To gather more detailed information, 5 test users were asked about their most important locations. A summary of the results for the evaluation metrics calculated for the 5 test users can be found in Table 3b. The algorithm detects 62% of the test users' locations of interest. One of the reasons that 38% of important locations listed by the users is not discovered is because, although the location is perceived as important, the user does not spend that much time there on a weekly basis compared to other locations. For example, a user might list the home of a family member he visits once a week but does not mention the train station he visits every day. Another explanation could be that the GPS signal is systematically turned off when visiting a certain location - deliberately or accidentally - which makes it very hard to discover that location. Most of the discovered addresses are close to the actual address, with only small mistakes in the house numbers.

All test users' homes were predicted correctly, even when including only 2 weeks of data. However, in this small test set no test users with two different home locations are included. Work location prediction is a little bit less accurate: 12% of work locations can not be discovered. This is mainly an issue when a user lists more than one work location. Also, for one of the users the previous work location was discovered. To improve work location predictions, including data like current and previous job on LinkedIn might be useful. However, the benefits and costs (collection, storage, privacy, etc.) of including this extra data will need to be taken into account.

Usefulness

To evaluate usefulness, we compared the locations discovered by our algorithm to the actual parcel delivery addresses of 305 users (Figure 4). *Useful recall* was 75%, meaning that 75% of the parcel delivery addresses was discovered by our algorithm, i.e. was in the user's top 4 locations. Furthermore, more than 50% of the delivery addresses was at a predicted home or work location. However, when considering the top 4 locations of each user, *useful precision* is only 41 %. All in all, it is quite a challenge to decide whether or not a location is useful as a delivery address based on the user's position history data alone. Therefore, previous delivery addresses should receive priority to serve as future delivery address candidates. For the moment, we will focus on home and work locations as the most important delivery address candidates.

Note that 19% of the delivery addresses were new locations, where the user was never seen before. No data about these locations was available in our dataset for the specific user and therefore these locations are impossible to predict from the position history data alone. This can occur when someone orders a parcel but asks to deliver it at a friend's location for example. This demonstrates the importance of keeping delivery data, as a valuable addition to position history data alone. Another explanation might be that this is a location where GPS tracking is systematically switched of, deliberately or not. Finally, in our case, some of these addresses are generated because of inaccurate data entry. When a parcel is delivered, the courier confirms delivery through his smartphone. The delivery address that will be stored in the database is automatically generated based on the courier's GPS position. However, when the courier confirms delivery a few 100 meters away from the actual delivery address and does not change the address manually, the address that is stored might differ from the actual address. Such findings are important to consider for further improvements of a company-tailored data science system.

Also for the 5 test users, the *useful recall* and *useful precision* are calculated (Table 3b) and show similar results. Almost all useful locations are discovered but only half of the discovered locations are useful.

[Figure 4 about here.]

Timeliness

Both the average home recall and precision rise when including users with more data until around respectively 50 and 30 days of data (Figure 5). However, the results must be interpreted with some caution, as the number of users decreases when the minimum amount of data increases (Figure 5c). Therefore, bootstrapping was performed to generate ranges around the average recall and precision with the minimum and maximum values of 1000 random samples with replacement. The range becomes wider with an increasing number of days, indicating more uncertainty and complicating conclusions. From 75 days onwards, ranges become too wide as less than 20 users remain and our analysis is therefore constrained to 75 days. Hence, with the current results we propose a 40-50 day window to keep the data but it would be interesting to repeat this experiment with more test users to have more reliable results.

Also from the 5 test users is seems that including more months of data benefits the discovery of locations of interest by the algorithm, up to a certain limit. Although performance was improved when including 2 months of data instead of 2 weeks, there was no difference in performance when including 5 months of data compared to 2 months. Including more data might improve results (up to a certain limit) but it will slow down computing time and increase storage costs. Moreover, for privacy reasons it is not desirable to store more customer data than needed.

[Figure 5 about here.]

Future location prediction

The first baseline model, based on historical frequencies achieved an average accuracy of 61% with a standard deviation of 17%. This standard deviation reflects the difference in predictability for the different users. The predictability is dependent on the amount of data available but also on the user's behavior. One user has a more stable location pattern than the other. To gain more insights in how this model is performing, a confusion matrix is constructed (Figure 6a). The model predicts the user will be at his home location in most of the cases, therefore the home location is predicted correctly for 76%, but the precision is low. The model performs less good in predicting the work locations.

The decision tree model performs a little bit better with an average accuracy of 63% and a standard deviation of 18%. From the confusion matrix (Figure 6b) we conclude that the model predicts less home locations correctly but the precision is slightly higher than the historical counts model. Also more work locations are predicted correctly.

[Figure 6 about here.]

The decision tree model seems to achieve a higher accuracy for users with more data available (Figure 7a). The accuracy is growing slightly until 70 days. Again, the amount of users with a high minimum of days with data is limited (Figure 7b) and therefore bootstrapping was performed to generate a range around the average accuracy. As we are again limited to 100 days, further research with more data is required to confirm conclusions.

[Figure 7 about here.]

The two models described above make location predictions one week ahead and are therefore mainly useful for long-term delivery planning. When the parcels to deliver are known in advance, the location predictions can serve as input for a route-optimization algorithm that minimizes both transportation costs and failed deliveries. This way, decision-making on route planning and hub selection can be improved. For the moment, only the user's home and work location are considered as these locations proved to be good candidates for delivery addresses. The model can easily be extended to include other addresses such as previous delivery addresses. On the other hand, for short-term planning, the next location prediction techniques described in the Literature overview are more suitable. Also, as the predictive power present in the historical location data alone is limited, external data such as Facebook friends and events, weather forecasts and calendar data, could be included to improve results. Of course, the same trade-off between the costs and benefits of including more data should be considered.

APPLICATION FOR LAST-MILE DELIVERY

We assume a scenario where the delivery service provider knows which customers to deliver one week in advance. In order to plan delivery to a customer two questions need to be answered: (i) *where* does the customer want to receive a delivery and (ii) *when* does the customer want to receive a delivery at a certain address? To answer the first question the location discovery and labeling is used to select potential delivery addresses candidates. Depending on the data available, a cascade of options can be considered: first of all, previous delivery addresses could be good candidates for future delivery addresses, because it is likely that the user wants to receive a parcel there. In this case, the delivery time can be based on previous delivery hours or on the presence probability predictions, which we will discuss in more detail next. If no previous delivery addresses are known for a certain user, the home and work location are the second best candidates for parcel delivery locations. Finally, if home and work location are not known or for some reason not available for delivery, other locations of interest can be considered.

The proposed location predictions can be used by delivery service providers in two different ways to improve delivery scheduling on an individual level. On the one hand, given a day/time slot, a user's most likely location can be predicted, answering the *where*-question. On the other hand, given a location, the user's presence probability can be predicted, answering the *when*-question.

The presence probabilities for a user's home and work location are represented in Figure 8 (a and b). This user seems to leave home around 7h-8h in the morning on weekdays and seems to return home after 20h. Only on Thursday this pattern differs and the user seems to leave home earlier and return later. The chances are higher to find this customer at home on Sundays rather than Saturdays but the probability is still only around 50%. The probability for being at home is highest during the night, but this is of course not a desirable delivery timing. The probability for being home is also high after 20h in the evening (and not on Thursdays). The customer arrives at work around 8h-9h in the morning and stays until 19h. The probability for being at this work location is higher on Monday, Tuesday and Wednesday compared to Thursday and Friday. Thursday is again a deviant day: the probability for being at this work location is lower in general but if the user is at this work location he seems to stay longer than on other days. The delivery service provider can plan an optimal route based on the weekly overviews of all customers that need to be delivered and suggest a location and time window to the customer that is convenient to both parties. In case the customer declines the suggestion a new route will need to be calculated. For this reason it is important to optimize the probability of acceptance by the customer.

Of course, in times of same-day delivery, not all deliveries will be known one week in advance. For short-term deliveries, the weekly schedule can be updated with the most recent information, such as *next* location prediction as described in the Literature Overview.

[Figure 8 about here.]

Next to individual location prediction, locations can also be predicted on the global level. Averaging the results over multiple users provides insights in home and work presence hours of a certain target audience in general. In Figure 8 (c and d) this was done for 30 users. Only users with at least three weeks of location data and a known work and home address (provided through the app) were selected. Although 30 users is too little to draw generalizing conclusions, we expect these users to be most likely at home after 20h in the evening until 7h-8h in the morning on weekdays. From 9h till 18h users are most likely not at home. During the weekend the probability for being home is more equal during the day but low in general. Most customers seem to arrive at work between 8h and 10h. Between 10 and 17h the probability for being at work is highest with a little downfall between 12h and 14h (this could be the lunch break). After 17h more and more customers leave work. As expected, the probability for being at work is low in the weekend. If one needs to deliver an arbitrary customer from this group without any individual information on presence hours, it would be best to deliver at his home location after 20h or at his work location on workdays between 10h and 12h or between 14h and 17h. Furthermore, global location prediction can be used to strengthen individual models.

DISCUSSION AND LIMITATIONS

The discovery and labeling of customer locations offer insights in the possible addresses to deliver, whereas the presence probabilities can support planning when to deliver at a certain address. With this information, failed deliveries can be reduced and delivery will become more customer-centered. The best results will be obtained when the information is used on the individual level and a user-tailored delivery service can be provided. However, we have shown that the location predictions can also provide useful insights at the group level. Users can be clustered based on socio-demographic, geographic or other characteristics and per group of users the average attendance hours could be predicted. If a new and unknown user can be positioned in one of the groups, information about the optimal delivery timing is already available. This way this study can also benefit delivery service providers without access to their individual user's location data.

The location prediction brings new opportunities for improved vehicle routing. The Vehicle Routing Problem (VHP) finds a set of routes for vehicles based at a depot, so that each of the customers that needs to be serviced is visited once, while minimizing the overall routing costs (Pillac, Gendreau, Guéret, & Medaglia, 2013). A number of variants of this problem have been studied but our application is most in line with the Vehicle Routing Problem with Time Windows (VRPTW), where each customer must be visited (delivered) within a certain time window. In our case however, the time window is not fixed but for each location the optimal time window can be predicted with a certain probability. Given these probabilistic time windows, the optimal route can be calculated. The different addresses and time windows to consider add a whole new level to the routing problem and would be an interesting avenue for future research. Moreover, with crowd sourcing finding its way to the field of last-mile delivery, the same techniques can be applied to the parcel pick-up side, in the end optimizing the whole process from pick-up to delivery. This would be a fit with the Vehicle Routing Problem with Pickup and Delivery (VRPPD).

Also for other applications inside or outside the field of transportation this methodology can be applied. At large scale, the weekly overviews offer new insights in citizen's transport behavior, which can be used to improve transportation infrastructure. Predicted home-work traffic can be used to optimize road design, public transport and traffic monitoring. Location prediction can support detailed mapping of crowded locations for every hour of the week, so appropriate measures can be taken for urban planning and transportation. Also for weekly energy demand forecasting this could be useful. At the individual level, weekly location predictions offer opportunities for location-based services such as long-term recommender systems (e.g. on Tuesday you could be reminded to wash your sports clothes for your weekly visit to the gym on Thursday) or advertising. Even smart energy appliances could use location predictions as input to learn an individual's behavior and improve energy efficiency in buildings.

Despite the promising outlook of this proposed approach, several limitations need to be considered and offer opportunities for future research. The proposed three-step approach offers a guiding framework to delivery service providers aiming to improve a customer-centered delivery service, but the individual techniques used can be further refined. For example, to improve presence probability predictions, other models as discussed in literature and combinations thereof might be considered. Furthermore, additional research on the optimal amount of data to include into the training of a prediction model could be investigated on a more elaborated dataset. Althought the total amount of users and the time span covered by the dataset used were rather extensive, the amount of active users was insufficient to draw substantiated conclusions. A future solution could be to supplement the historical GPS locations with external data, such as social media data. The advantages and disadvantages of including more data should be considered.

Before putting this approach into practice, further field testing is required to select the appropriate, case-specific training data and algorithms. Real-life implementation provides useful feedback in terms of how often and in which way the location predictions are used by schedulers. This feedback can be used by delivery service providers to improve the method, tailored to their own needs. Finally, it is important to consider legal issues with respect to data privacy and security. Customer locations are very sensitive data and need to be collected, stored an processed in a secure and privacy-friendly way. The user must be aware of which data is collected, how it is used and who has access to it, being able to edit or delete data at all times. Not for all situations this location data is at hand, opening doors for research on this matter using other kinds of (less sensitive) data.

CONCLUSION

The goal of this research was to predict the future locations of individuals, based on mobile location data, in order to support delivery planning for last-mile delivery companies. A three-step approach was proposed that discovers a user's locations of interest through density based clustering, labels the locations according to temporal assumptions and finally predicts locations in the future.

According to the results, the proposed approach can detect and label most of the user's important locations correctly and shows a relatively exact address except for the house numbers. Including more days of data improves the discovery and labeling of locations of interest up to a certain limit. Although the data available is insufficient to draw substantiated conclusions on this matter, it seems optimal to include around 40 days. A simple decision tree model can predict when a user is at home or at work with an average accuracy of 63%. The accuracy again is higher for users with more data, which implies that including more data improves results. However, the data available was too limited to draw substantiated conclusions concerning the optimal amount of data to include.

Our method can improve decision making for last mile delivery services aiming to minimize transportation costs while optimizing customer service. In most of the cases the address to deliver was a predicted home or work location, indicating that our method can offer helpful insights as to *where* to deliver to a customer. On the other hand, the presence probabilities per location can support planning *when* to deliver at a certain address. This way, our approach will help delivery service providers to gain insights into their customers' habits and thus their optimal delivery time and location, ultimately reducing the rate of failed deliveries. For customers, this implies that they need to undertake a minimum effort in order to receive a parcel which will eventually boost customer satisfaction. Finally, more efficient parcel deliveries are not only beneficial for delivery service providers and their customers but will also decrease the burden of transportation on the environment. Several interesting opportunities for future research are possible in the fields of vehicle routing, transportation infrastructure and even energy demand forecasting.

References

Arnold, F., Cardenas, I., Sörensen, K., & Dewulf, W. (2016). Simulation of different scenarios for b2c e-commerce distribution in antwerp. In *Proceedings of the vref conference on urban freight 2016: Plan for the future-sharing urban space: 17-19 october, 2016, gothenburg, sweden* (pp. 1–11).

Asahara, A., Maruyama, K., Sato, A., & Seto, K. (2011). Pedestrian-movement prediction based on mixed markov-chain model. In *Proceedings of the 19th acm sigspatial international conference on advances in geographic information systems* (pp. 25–33).

Ashbrook, D., & Starner, T. (2003). Using gps to learn significant locations and predict movement across multiple users. *Personal and Ubiquitous Computing*, *7*(5), 275–286.

Bramer, M. (2007). *Principles of data mining* (Vol. 180). Springer.

Comeos. (2018). *E-commerce belgium 2018.* Brochure. Retrieved 2018-09-11, from `https://static.comeos.be/4._E-commerce_full_1.pdf`

Dash, M., Nguyen, H. L., Hong, C., Yap, G. E., Nguyen, M. N., Li, X., ... others (2014). Home and work place prediction for urban planning using mobile network data. In *2014 ieee 15th international conference on mobile data management* (Vol. 2, pp. 37–42).

Do, T. M. T., & Gatica-Perez, D. (2014). The places of our lives: Visiting patterns and automatic labeling from longitudinal smartphone data. *IEEE Transactions on Mobile Computing*, *13*(3), 638–648.

Eftelioglu, E. (2015). Geometric median. In S. Shekhar, H. Xiong, & X. Zhou (Eds.), *Encyclopedia of gis* (pp. 1–4). Cham: Springer International Publishing. Retrieved from `https://doi.org/10.1007/978-3-319-23519-6_1555-1` doi: 10.1007/978-3-319-23519-6_1555-1

Ester, M., Kriegel, H.-P., Sander, J., Xu, X., et al. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In *Kdd* (Vol. 96, pp. 226–231).

Gambs, S., Killijian, M.-O., & del Prado Cortez, M. N. (2012). Next place prediction using mobility markov chains. In *Proceedings of the first workshop on measurement, privacy, and mobility* (p. 3).

Gevaers, R., Van de Voorde, E., & Vanelslander, T. (2011). Characteristics and typology of last-mile logistics from an innovation perspective in an urban context. *City Distribution and Urban Freight Transport: Multiple Perspectives, Edward Elgar Publishing*, 56–71.

Gonzalez, M. C., Hidalgo, C. A., & Barabasi, A.-L. (2008). Understanding individual human mobility patterns. *Nature*, *453*(7196), 779–782.

Huang, C., Jia-ching Ying, J., & Tseng, V. S. (2012). Mining users' behaviors and environments for semantic place prediction. In *Mobile data challenge 2012 (by nokia) workshop; june 18-19, 2012; newcastle, uk.*

Huang, Q., & Wong, D. W. S. (2016). Activity patterns, socioeconomic status and urban spatial structure: what can social media data tell us? *International Journal of Geographical Information Science*, *30*(9), 1873-1898. Retrieved from `https://doi.org/10.1080/13658816.2016.1145225` doi: 10.1080/13658816.2016.1145225

Ingrid, E. B. (2011). Predicting future locations and arrival times of individuals.

Kang, J. H., Welbourne, W., Stewart, B., & Borriello, G. (2004). Extracting places from traces of locations. In *Proceedings of the 2nd acm international workshop on wireless mobile applications and services on wlan hotspots* (pp. 110–118).

Krumm, J., & Brush, A. B. (2011). Learning time-based presence probabilities. In *International conference on pervasive computing* (pp. 79–96).

Lee, S., Lee, K. C., & Cho, H. (2010). A dynamic bayesian network approach to location prediction in ubiquitous computing environments. In *International conference on advances in information technology* (pp. 73–82).

Mathew, W., Raposo, R., & Martins, B. (2012). Predicting future locations with hidden markov models. In *Proceedings of the 2012 acm conference on ubiquitous computing* (pp. 911–918).

Monreale, A., Pinelli, F., Trasarti, R., & Giannotti, F. (2009). Wherenext: a location predictor on trajectory pattern mining. In *Proceedings of the 15th acm sigkdd international* *conference on knowledge discovery and data mining* (pp. 637–646).

Montoliu, R., & Gatica-Perez, D. (2010). Discovering human places of interest from multimodal mobile phone data. In *Proceedings of the 9th international conference on mobile and ubiquitous multimedia* (p. 12).

Morzy, M. (2007). Mining frequent trajectories of moving objects for location prediction. *Machine Learning and Data Mining in Pattern Recognition*, 667–680.

Pan, S., Giannikas, V., Han, Y., Grover-Silva, E., Qiao, B., et al. (2017). Using customer-related data to enhance e-grocery home delivery. *Industrial Management and Data Systems*.

Parcify. (2016). *What is Parcify.* Retrieved 2016-12-16, from `https://www.parcify.com/what-is-parcify/`

Petzold, J., Bagci, F., Trumler, W., & Ungerer, T. (2003). The state predictor method for context prediction. In *Adjunct proceedings fifth international conference on ubiquitous computing* (pp. 135–147).

Petzold, J., Pietzowski, A., Bagci, F., Trumler, W., & Ungerer, T. (2005). Prediction of indoor movements using bayesian networks. In *International symposium on location-and context-awareness* (pp. 211–222).

Pillac, V., Gendreau, M., Guéret, C., & Medaglia, A. L. (2013). A review of dynamic vehicle routing problems. *European Journal of Operational Research*, *225*(1), 1–11.

Ryan, C., & Brown, K. N. (2012). Occupant location prediction using association rule mining. In *Workshop on ai problems and approaches for intelligent environments* (p. 27).

Sadilek, A., & Krumm, J. (2012). Far out: Predicting long-term human mobility. In *Aaai.*

Statista. (2016). *Global retail e-commerce sales 2014-2020.* Retrieved 2016-12-15, from `https://www.statista.com/statistics/379046/worldwide-retail-e-commerce-sales/`

Van Duin, J., De Goffau, W., Wiegmans, B., Tavasszy, L., & Saes, M. (2016). Improving home delivery efficiency by using principles of address intelligence for b2c deliveries. *Transportation Research Procedia*, *12*, 14–25.

Vintan, L., Gellert, A., Petzold, J., & Ungerer, T. (2004). Person movement prediction using neural networks. In *First workshop on modeling and retrieval of context* (Vol. 114, pp.

1–12).

Wang, X., & Hamilton, H. (2009). Using clustering methods in geospatial information systems. In *Geoinformatics 2008 and joint conference on gis and built environment: Advanced spatial data models and analyses* (pp. 71461N–71461N).

Ye, Y., Zheng, Y., Chen, Y., Feng, J., & Xie, X. (2009). Mining individual life pattern based on location history. In *2009 tenth international conference on mobile data management: Systems, services and middleware* (pp. 1–10).

Zhou, C., Frankowski, D., Ludford, P., Shekhar, S., & Terveen, L. (2007). Discovering personally meaningful places: An interactive clustering approach. *ACM Transactions on Information Systems (TOIS)*, *25*(3), 12.

Table 1: Literature overview of the application problems relevant to this study, summarizing different techniques and application domains.

Application problem	Techniques	Papers	Application domains
Discovering significant locations	K-means clustering	Montoliu & Gatica-Perez, 2010; Ashbrook & Starner, 2003	Mobile applications
	Time-based clustering	Kang et al., 2004	Mobile applications
	Density-based clustering	Ye et al., 2009	Mobile applications
	Grid-based clustering	Do & Gatica-Perez, 2014	Mobile applications
Semantic location labeling	Random forest for multi-class classification	Do & Gatica-Perez, 2014	Mobile applications
	Inactivity and temporal features	Dash et al., 2014	Home and work location prediction for urban planning
Next location prediction	Markov chain	Ashbrook & Starner, 2003	Mobile applications
	Mobility Markov Chain	Gambs et al., 2012	Mobile applications
	Hidden Markov Models	Mathew et al., 2012	Mobile applications
	Mixed Markov Model	Asahara et al., 2011	Shopping and selling services
	Association rule mining	Ryan & Brown, 2012	Energy efficiency in buildings
	Movement rules extraction	Morzy, 2007; Monreale et al., 2009	Mobile applications
	Bayesian networks	Lee et al., 2010; Petzold et al., 2005	Mobile applications, smart office environment
	Neural networks	Vintan et al., 2004	Mobile applications
	State predictors	Petzold et al., 2003	Mobile applications
Future location prediction	Prediction-by-Partial Match algorithm	Ingrid, 2011	Mobile applications
	Principal Component Analysis	Sadilek & Krumm, 2012	Mobile applications, urban planning, spread of disease, demand for electricity, etc.
	Probabilistic schedule	Krumm & Brush, 2011	Energy efficiency

Table 2: Features of the final position history dataset

Feature	Description
userId	Unique code representing a user
Latitude	The latitude coordinate of the user's location
Longitude	The longitude coordinate of the user's location
GPS inaccuracy	The error margin of the user's GPS position, specified in meters
Start time	Date and time the user entered the location
Time delta	Time the user stayed at the location
End time	Date and time the user left the location

Table 3: Results for the evaluation metrics for the user addresses evaluation (a) and for the 5 test users when using 5 months, 2 months or 2 weeks of data (b).

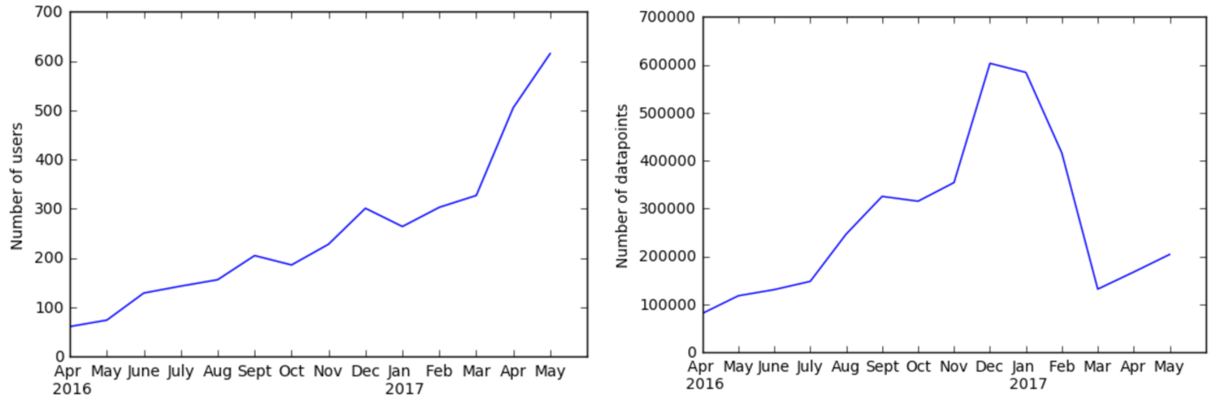
(a) User address evaluation

Evaluation metric	Result
Home recall (%)	80
Home precision (%)	76
Work recall (%)	62
Work precision (%)	30
Useful recall (%)	75
Useful precision (%)	41

(b) Test users evaluation

Evaluation metric	5 Months	2 Months	2 Weeks
Recall (%)	62	62	52
Precision (%)	63	63	56
Useful recall (%)	90	90	82
Useful precision (%)	53	53	45
Home recall (%)	100	100	100
Work recall (%)	88	88	63

Figure 1: The number of active users in the dataset is increasing over time (a). The number of datapoints is increasing until December 2016, after which the data collection system gradually switched to more efficient data storage (b).



(a) Active users per month

(b) Datapoints per month

Figure 2: Overview of the process followed in this work. From the raw GPS data we start by discovering and subsequently labeling locations of interest. Next, prediction methods are used to predict at which location a user will most likely be in the future. The final step is the evaluation of the approach.

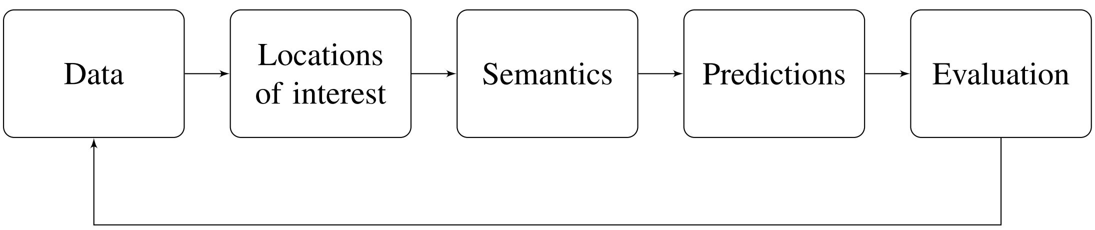
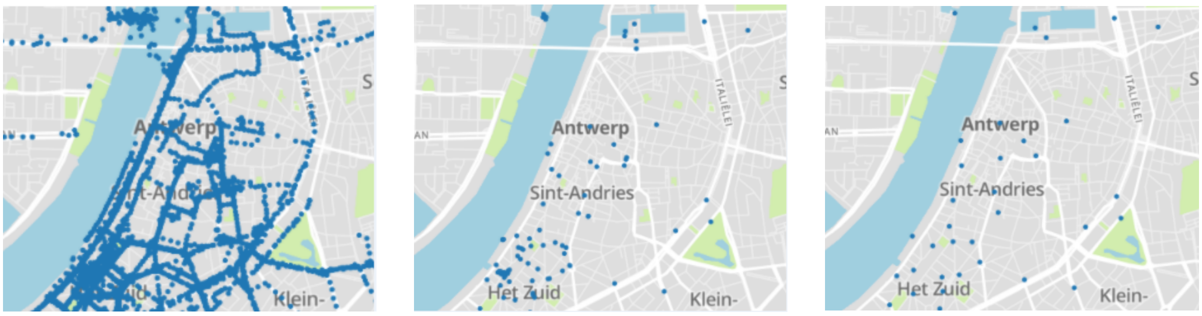


Figure 3: An overview of our approach to discover locations of interest. From all the historical location points of a user (a), only those locations where a user stayed for more than 5 minutes are selected (b). Next, these locations are clustered into stay regions (c).



(a) Location points **(b)** Stay points **(c)** Stay regions

Figure 4: Out of the 511 actual delivery addresses, 383 addresses were labeled by our algorithm as locations of interest, 29 addresses were found in the user's location history (visited location) but not labeled as a location of interest and finally, 99 addresses were new locations, where the user was never seen before.

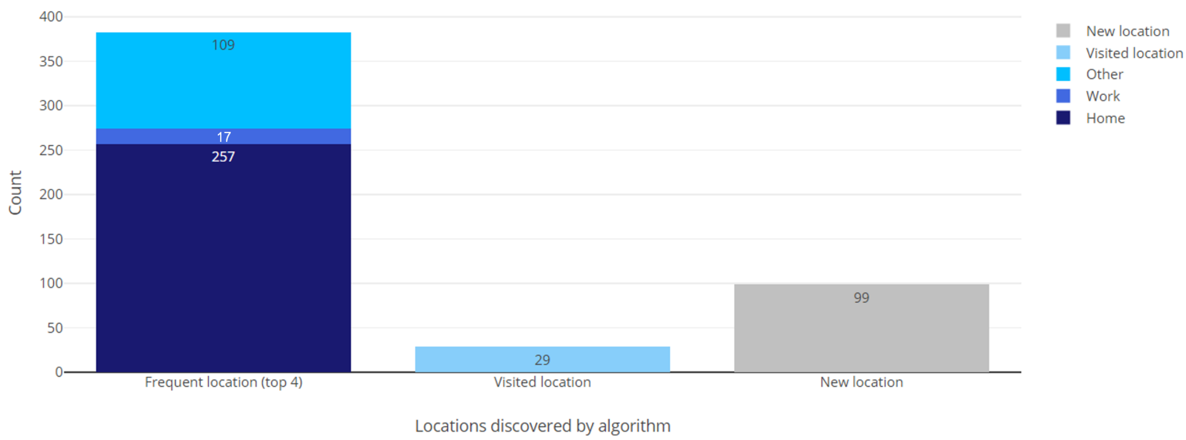
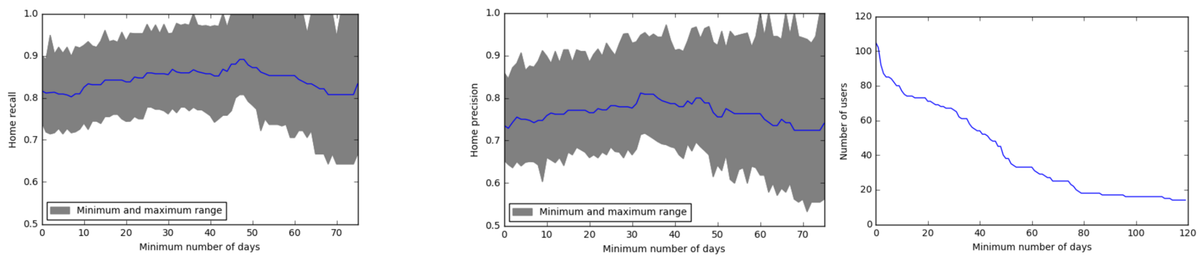
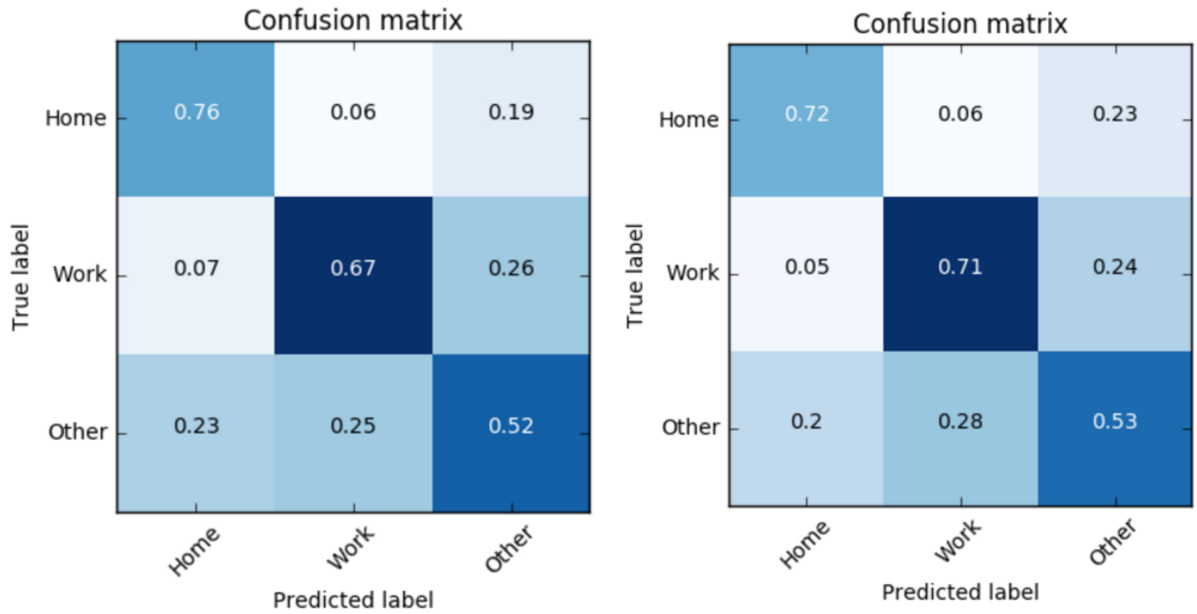


Figure 5: Having more days with data available seems to improve home recall (a) and precision (b) up to a certain limit. However, some caution is required interpreting the results, as the number of users with a high minimum number of days might be insufficient to draw substantiated conclusions (c).



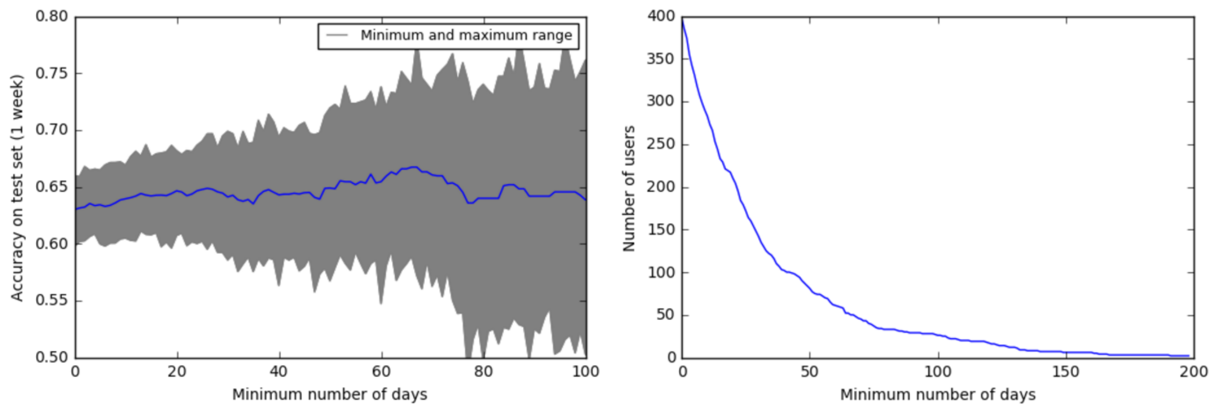
(a) Home recall **(b)** Home precision **(c)** Number of users

Figure 6: (a) The historical counts model predicts the user will be at his home location with low precision and performs poorly at predicting when a user will be at the work location. (b) The decision tree model performs better at predicting when a user will be at his work location but slightly less for his home location.



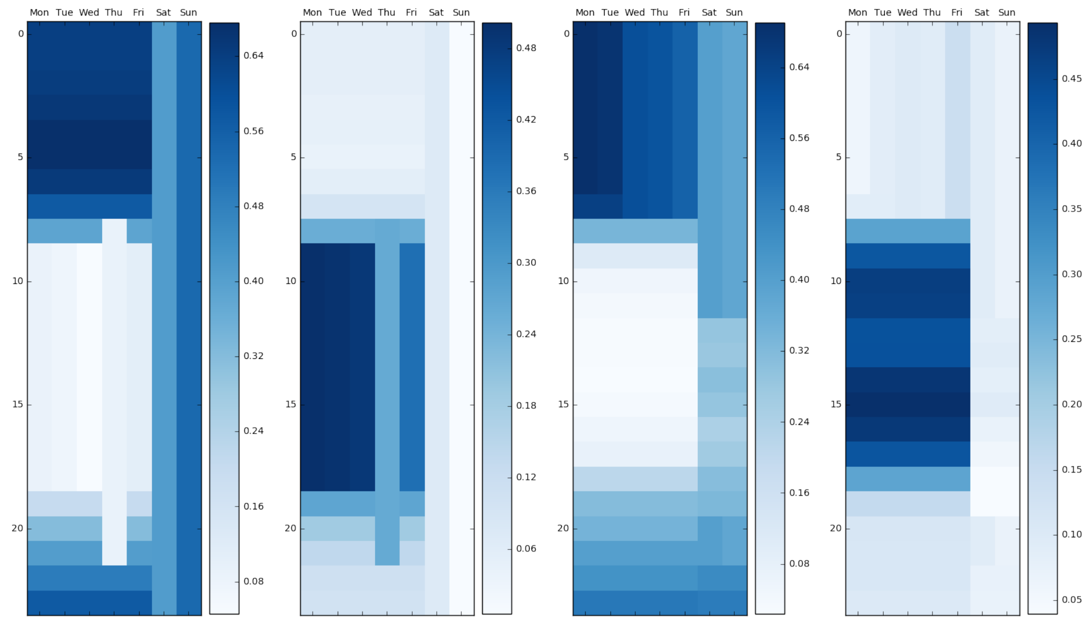
(a) Confusion matrix for the historical counts model **(b)** Confusion matrix for the decision tree model

Figure 7: Having more days with data available seems to improve the accuracy of the location predictions of the decision tree model (a). However, some caution is required interpreting the results, as the number of users with a high minimum number of days might be insufficient to draw substantiated conclusions (b).



(a) Accuracy on test set **(b)** Number of users

Figure 8: For delivery scheduling, a weekly overview of the users' presence probability for different locations is represented as a colormap. A darker color reflects a higher probability. This example shows an individual user's home (a) and work (b) location presence probability and the average presence probability for home (c) and work (d) location over 30 users



(a) Home location (individual)

(b) Work location (individual)

(c) Home location (average)

(d) Work location (average)