UNITED KINGDOM · CHINA · MALAYSIA

Amirian, Pouria and Basiri, Anahid and Morley, Jeremy (2016) Predictive analytics for enhancing travel time estimation in navigation apps of Apple, Google, and Microsoft. In: 9th ACM SIGSPATIAL International Workshop on Computational Transportation Science, 31 October - 3 November 2016, Burlingame, California, USA.

Predictive Analytics for Enhancing Travel Time Estimation in Navigation Apps of Apple, Google, and Microsoft

Pouria Amirian
Ordnance Survey of Great Britain,
Products and Innovation,
Southampton, SO16 0AS, UK
+44(0) 7775221531
Pouria.Amirian@os.uk

Anahid Basiri
The University of Nottingham
Nottingham Geospatial Institute (NGI)
Nottingham, NG7 2TU, UK
+44(0)1158467850
Anahid.Basiri@nottingham.ac.uk

Jeremy Morley
Ordnance Survey of Great Britain
Products and Innovation,
Southampton, SO16 0AS, UK
+44(0)2380055168
Jeremy.Morley@os.uk

ABSTRACT
The explosive growth of the location-enabled devices coupled with the increasing use of Internet services has led to an increasing awareness of the importance and usage of geospatial information in many applications. The mobile navigation apps (often called "Maps"), use a variety of available data sources to calculate and predict the travel time for different modes. This paper evaluates the pedestrian mode of Maps apps in three major smartphone operating systems (Android, iOS and Windows Phone). We will demonstrate that the Maps apps on iOS, Android and Windows Phone in pedestrian mode, predict travel time without learning from the individual's movement profile. Then, we will exemplify that those apps suffer from a specific data quality issue (the absence of information about location and type of pedestrian crossings). Finally, we will illustrate learning from movement profile of individuals using predictive analytics models to improve the accuracy of travel time estimation for each user (personalization).

Keywords
Predictive Analytics; Navigation; Movement Profile; Pedestrian; Location-Based Services; Personalization

1. INTRODUCTION
Nowadays, increasing use of location-enabled devices has led to an increasing awareness of the importance and usage of geospatial information in many applications. Google recently reported that over half of trillions of searches on Google.com in recent years happened on smartphones. Nearly one-third of all mobile searches on Google.com are related to location. More importantly, location-related mobile searches are growing 50% faster than mobile searches in general. In the navigation context, use of smartphones and their navigation apps, have been replacing satnav devices for car navigation [2]. In general, navigation apps are more efficient for they can easily access to real-time traffic and are able to recommend alternative routes based on the current situation of the road network. They are economically more appealing than satnavs and from the user experience, for the users there is no need to buy a single purpose device just for navigation, which needs content or software updates once in a while. In fact, all major smartphone platforms (Android, iOS, and Windows Phone) provide at least one native mobile app for navigation (often called "Maps") that pre-

loaded by device manufacturers. The Maps apps, use a variety of

available data sources to calculate and predict the travel time as well as several options for routing in public transportation, car or pedestrian modes. For car and public transportation routing, the apps use available data from multitude of sources like official speed limits of routes, real-time speeds derived from transportation sensors in routes, historical average speed data over certain time periods, actual travel times from previous users, and real-time traffic and incidents information. In order to make higher accuracy prediction most of the Maps apps, act not only as a consumer of the real-time data but also as real-time data generators; the Maps app continuously send information about the location of the devices. Then a huge amount of location data from a large number of devices is used to determine various real-time information about current status of road networks. In addition, the collected data are utilized as historical data for tuning the prediction algorithms for routing and estimation of travel time. That is why, most of the Maps apps, can provide more accurate predictions when connected to the Internet. However, for pedestrian mode, the main (and in most cases the only) source of estimation of travel time is the pedestrian road network. Most of the Maps apps tend to use just a formula with a single set of parameters for all pedestrians to calculate the travel time without considering individual users' characteristics. Using a formula for all users is not terribly a bad idea. A pedestrian tends to have inertia in her movement characteristics and usually doesn't change her walking speed frequently. In addition, real-time traffic and incidents in the car and public transportation networks have little impact on individuals' walking speed. The above reasons seem to be rational enough for using just the pedestrian road network for estimation of travel time for pedestrians. In this case, as the user starts walking, the apps provide an estimated travel time based on a fixed walking speed and length of the route. The travel time is updated at fixed time intervals (usually, a few seconds depending on the app). The estimated travel time mainly is calculated based on the current location of the user (derived from smartphone's GPS) and its distance to the destination (we call this approach *"naïve"* approach). In other words, historical data of the owners of devices are not considered for the calculation of the estimation of travel time; that is why for all people at the same location and with the same destination, Maps apps (of same mobile platform) estimate identical values for the travel time.

While highly intelligent and sophisticated algorithms are being used for different purposes like providing personalized real-time advertisement and recommendation, no clever algorithm or approach have been used for pedestrian navigation. In pedestrian mode, the Maps apps do not collect data for each user to learn about the movement characteristics of each individual. In other words, Maps apps (in pedestrian mode), on smartphones are not smart enough to utilize the valuable personal source of data for providing more accurate personalized navigation services. The current naïve approach of predicting travel time results in the less accurate estimation of duration of travel and as a result, it leads to anxiety for pedestrians especially in multi-modal travels. It was reported that users of pedestrian navigation guidance

sometimes feel anxiety because of discrepancy between the estimated and actual time of arrival to destination [8]. The naïve approach is also problematic from battery consumption point of view. The current naïve approach of navigation apps needs continuous receiving signals from several GPS satellites. Today's smartphones achieve their long lasting battery life largely because they can aggressively and quickly enter into and exit from sleep states. Use of GPS prevents this clever way of saving battery life. Although this is not a problem for using smartphones in a car, for pedestrians it is a serious problem. This paper tries to introduce an approach for using machine learning techniques to improve the estimation of travel time for each pedestrian. Following are the contributions of the paper:

- We evaluate and discuss the estimation accuracy of travel time in pedestrian mode in three major native apps for iOS, Android and Windows Phone (section 3).

- We illustrate the use of predictive analytics (supervised machine learning) in improving the accuracy of estimated travel time for pedestrian by learning from individual movement profile data (section 4).

The paper illustrates applying predictive analytics on data which can be easily collected for pedestrians using their smartphones in order to improve the accuracy of estimated travel time and provide personalized services. In this context, surprisingly none of the major routing and navigation apps for pedestrian mode provide such personalized experience for their users.

2. RELATED WORK

Providing personalized routing experience (mostly for drivers) is subject of several research projects. Delling et al. implemented a framework for generating personalized driving directions by automatically analyzing GPS traces [6]. Via examining routes from GPS logs, Letchner et al. [11] found that drivers took the fastest route, as given by a commercial routing engine, only 35% of the time. Chang et al. [4], developed a personalized router for drivers using trajectory mining technique to select routes that were most familiar to the driver. Personalization using Landmark-based pedestrian navigation has been discussed in [3]. The authors in Going My Way [5] proposed a personalized route planner for landmark-based pedestrian navigation. The system can identify the landmarks automatically from the personal historical GPS log data and provide navigational instructions based on the landmarks rather than street names. Since the Going My Way system learns from data collected from user's navigation history, the system would have no assistance for the user for places the user has never been to. Amirian et al [1] designed and implemented a system for providing personalized pedestrian navigation based on landmarks for tourists. In their work, navigational instructions were personalized for each user based on her movement profile.

The work presented in this paper is focused on the predictive analytics for estimation of travel time at the individual level for providing personalized travel time estimation rather than personalization of routing or navigational instructions.

## 3. ESTIMATION OF TRAVEL TIME IN PEDESTRIAN MODE

Pedestrian navigation in all native Maps apps of major smartphone platforms uses the naïve approach. After setting the destination (and starting point which usually is the current location of the user) and starting the navigation process, the apps, use GPS periodically (every few seconds) to locate the user and calculate the distance between current location and the destination. Based on these calculations the apps update estimated

travel time. In other words, for all user at the same location, the duration of travel to the same destination is identical in existing Maps apps (of the same platform). Despite the fact that available personal data can be used to provide more realistic and personalized travel time, it hasn't been used in any major navigation apps in all smartphone operating systems. By collecting individuals' movement profile data, it is possible to use modern predictive analytics methods to provide personalized pedestrian navigation services. In this research, we will illustrate the use of movement profile data to improve the predicted travel time for pedestrians. In addition, with approach of this research GPS sensor need to be used less frequently and therefore improve the battery life of the smartphones.

3.1 Accuracy of Prediction of Travel Time

We conducted an experiment to evaluate the three native Maps apps from major smartphone operating systems; Android, iOS, and Windows Phone. 39 people (21 men and 18 women) with an average age of 35.6 years participated in the experiment. They walked 48 different routes with an average length of 2.8 km (min= 0.8 km and max= 4.5 km) and each route was traveled at least 5 times in Oxford, UK. All routes were navigated with all of the three Maps apps. We used the default Maps app in iOS 9.3.1 and 9.3.2, Windows Phone 8.1 and 10 and Android 4.4.2, 5.1.2 and 6.0.1. Figure 1 shows one of the routes and its estimated travel time in three different apps for the same route (note the difference in time).
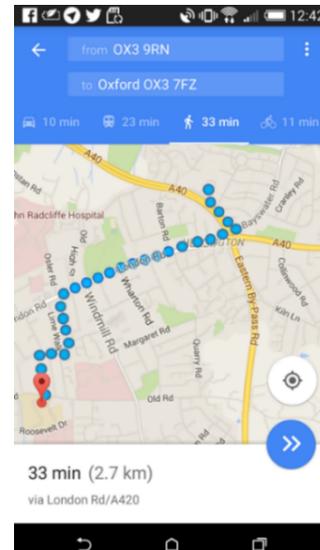
(a) Android. Estimated travel time 33 minutes.
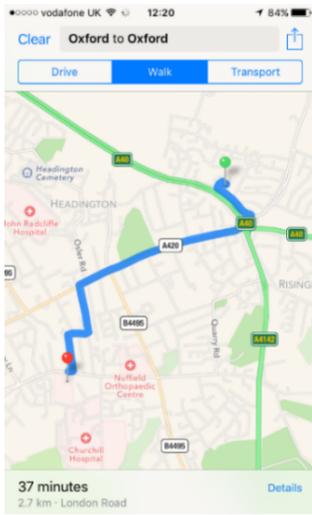
(b) iOS. Estimated travel time 37 minutes.
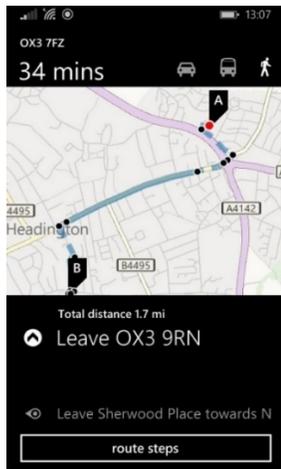

(c) Windows Phone. Estimated travel time 34 minutes.

Figure 1. A route of the experiment in three different Maps apps. Actual travel time is about 26-29 minutes.

In addition to estimated travel time, we asked all participants to measure the actual time of travel using a stopwatch or stopwatch apps then we calculated approximation error or relative error for each travel as $e = |t_{actual} - t_{predicted}|/t_{actual}$.

Figure 2,3 and 4 shows the relative error for each app, along with average of relative error and correlation between errors and length of routes (average error is indicated by horizontal red dotted line).

In all three apps, the length of the route is highly correlated with a relative error in prediction. In other words, as the length of route increases, the relative error linearly grows. In 83% of routes, the "Maps" app on iOS has the largest relative error in estimation of travel time in comparison with Android and Windows Phone apps. Android Maps was better (less relative error) in the estimation of travel time than Windows Phone Maps in 67% of routes. The estimations mostly were greater than the actual travel time. In just 4 routes (out of 48), the estimation of all apps were less than the actual travel time. In all the mentioned 4 routes, the routes for cross more than 3 highways. In other words, the

participants in those routes had to wait until the pedestrian lights turned to green.
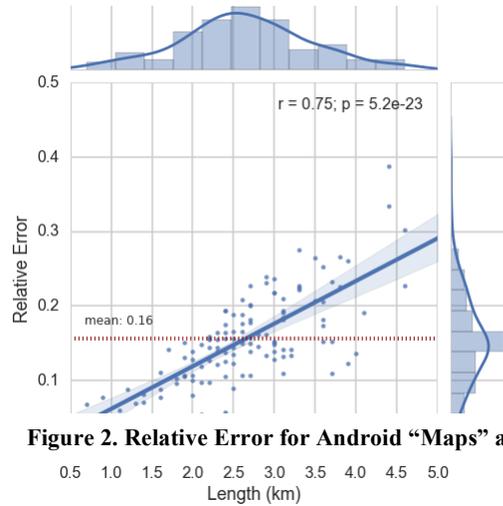
Figure 2. Relative Error for Android "Maps" app.
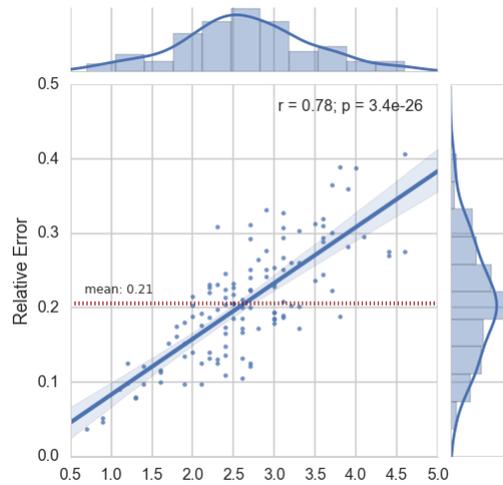
**Figure 3. Relative Error for iOS "Maps" app.**
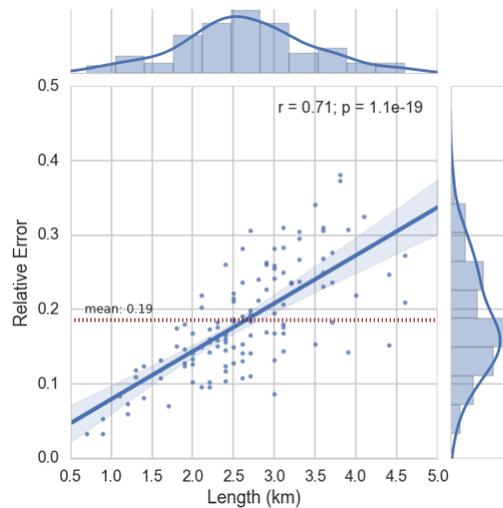
Figure 4. Relative Error in Windows Phone "Maps" app.

The highway code in the UK includes seven different types of pedestrian crossings each with various types of rules for pedestrians and drivers. It seems that a part of the high rate of error in prediction of travel time in pedestrian navigation is that the algorithms for routing in pedestrian mode do not differentiate between various types of the pedestrian crossings. With further exploration, we discovered that the algorithms do not use the location of pedestrian crossing in their calculations (seemingly due to lack of information at least in the case study area of this research, Oxford, UK). With further examination, we found that this is also the case for the online mapping services from Google and Microsoft. As an example, figure 5, shows a route in maps.google.com for crossing a highway to reach to a bus stop (as the initial stage of a multi-modal travel). Based on the suggested route it takes 2 minutes (146 meters) to reach the bus stop. Apparently, maps.google.com and Android Maps (figure 5), assumed that a pedestrian can cross the highway anywhere (which is not the case in real-world).
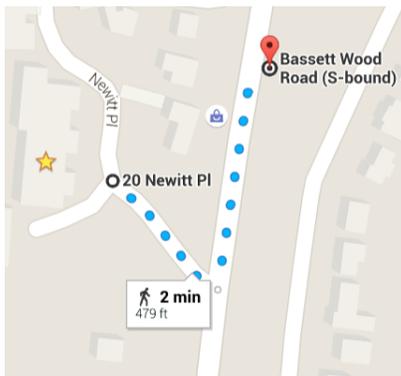
Figure 5. Suggested route for crossing the A33 road. It is predicted as 2 minutes walking distance.

Since pedestrians must pass the highway using a pedestrian crossing, they need to walk at least 3 times more than the suggested distance (figure 6). In contrast to what is suggested, it takes a pedestrian 482 meters to cross the highway. Regarding the travel time, if the pedestrian crossing was of type Zebra, that route would take 6 minutes; however, since the pedestrian crossing is of type "Puffin" it takes up to 8 minutes; since the pedestrians have to for cars. The large difference between estimated and actual travel time can lead to hours of delays in a multi-modal travel (if the user misses her train, because he missed the bus for example). In addition, safe crossing places is a very important characteristic for assessing the walkability of the pedestrian route [10] and an environment's walkability has a major impact of walking speed of individuals [7].

As it illustrated, neither online mapping services nor Maps apps consider walkability of pedestrian routes (at least for case study area of this research). Interestingly the location and type of pedestrian crossing are mostly available through OpenStreetMap project. In addition, with the availability of street view data and high accuracy machine learning algorithms for image recognition, it is possible to detect the location and type of pedestrian crossing automatically and improve the accuracy of pedestrian navigation. Nevertheless, this issue is related to data quality issue (incompleteness). Another part of the high rate of error in prediction of travel time is the lack of personalization and adaptation capabilities. In other words, neither of the native Maps apps, learn from movement profile of the user, weather condition,

and personal data in order to improve the accuracy of prediction of travel time for each user. Instead, they just simply recalculate the distance and travel time between current and the destination locations.
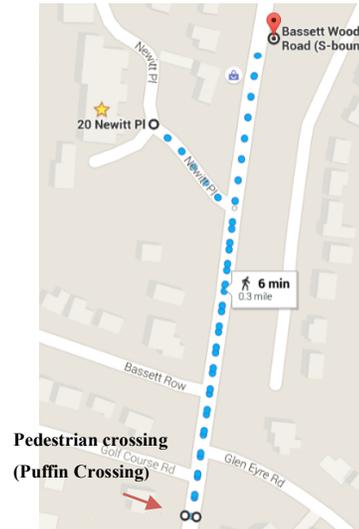
Figure 6. Actual route for crossing the A33 road (7-8 minutes of walking distance instead of 2 minutes).

## 4. PREDICTIVE ANALYTICS

4.1 Dataset

In the mentioned experiment, we collected some other attributes (or features in machine learning terminology) for each travel and participant in addition to the actual time of travel and estimated travel time. Interestingly the dataset of the experiment illustrates the changes of movement behavior based on different attributes. Most of the participants had different travel time for the same routes (and directions) at a different time in same days or in different days. It seems that combinations of environmental and temporal factors cause variable walking speed and thus differentiate between travel time for the same individual for the same route. Apart from environmental and temporal factors, differences between various travel time of various participants are highly correlated with individuals walking characteristics such as average walking speed and the total length of travel during the same day. The dataset for the predictive analytics contains attributes for temporal and environmental measurements about a certain route for certain person. The dataset composed of four kinds of features:

- Personal features such as age, gender, sum of length of journey for current day,
- Temporal and environmental features such as time of day, weekday and weather conditions,
- Geospatial features of the route such as length and change of elevation within the route,
- Travel time features. actual travel time and estimated travel time (estimated by Android's Maps app).

In the dataset of the experiment, all attributes were recorded by each participant except the sum of the length of the journey for current day and change of elevation within the route. For the length of the journey for current day attribute, we utilized digital pedometer sensors in smartphones. For each route, the length of

the journey of the current day is the number of steps that the user has taken before beginning walking the route. Also for each route we calculated the change of elevation attribute (elevation gain, loss, and absolute change of elevation for each route), using the Elevation API of Google Maps. Given the sensitivity of detailed movement data of individuals in this research, the dataset contains no raw movement data of individuals. While the anonymity of movement data is hard to achieve [14], it contains patterns that can identify individuals. In fact, De Montjoye et al. [13] studied fifteen months of human mobility data and found that four spatiotemporal points are enough to uniquely identify 95% of the individuals.

4.2 Predictive Analytics with All Features

In order to make a personalized prediction for each user, in this research, we trained several predictive models for predicting a correction value based on the environmental, temporal and personal factors. The predicted correction value needs to be added to the estimated value (from Apps) in order to personalize the estimated value of travel time for each user. Following figure shows the prediction accuracy using R^2 (coefficient of determination) for each model.
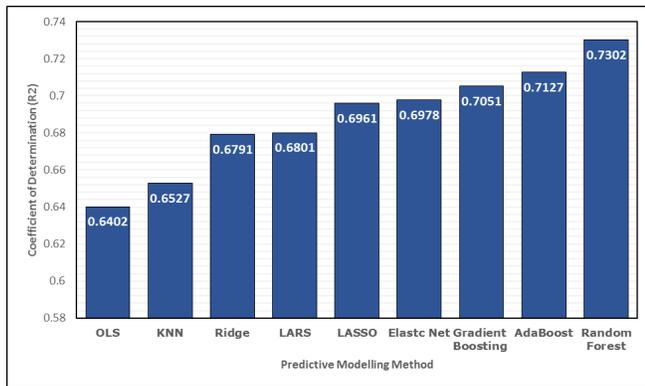

Figure 7. Prediction accuracy of various machine learning methods for all features. The prediction accuracy is calculated based on R^2 (out of sample accuracy) with cross validation.

As it illustrated in figure 7, even using simple methods like OLS produces a good result. Since travel time is a linear function of the length of the route, it was expected that the target variable (correction of travel time) could be determined by a linear combination of a subset of potential features. This is illustrated in figure 7 where OLS and penalized linear regression methods produce high accuracy prediction. The penalized linear regression methods, such as Ridge, LASSO, LARS, and Elastic Net generally produce better predictions than OLS solution through a better compromise between bias and variance. As it is shown in figure 7, ensemble learning methods like Gradient Boosting, AdaBoost and Random Forest yield most accurate predictions. Figure 8, shows the relative importance of features in the Random Forest method.

The feature importance is an estimation of prediction strength of each feature[9]. As it expected the length of the route is the most important feature for prediction of the correction value. Weather condition and Total steps (sum of the length of the journey for current day) are two other important features. As it illustrated in figure 8, gender is more important than the age feature. However, this might be due to sample bias in the experiment. In other words, in our experiment the age of participants was between 28 and 38 years (and the average was 35.6) which includes a specific

age range and definitely is not representative of the population which uses the smartphones. In the dataset, time of day is highly correlated with total steps and change of elevation has a high value of correlation with the length of the route. This might be the reason for the low importance of the change of elevation and time of day features. With correlated features, strong features can end up with low importance values. This is important especially with regards to the change of elevation since the correlation between length of route and change of elevation in the dataset is 0.7832 which means these two features are almost collinear.
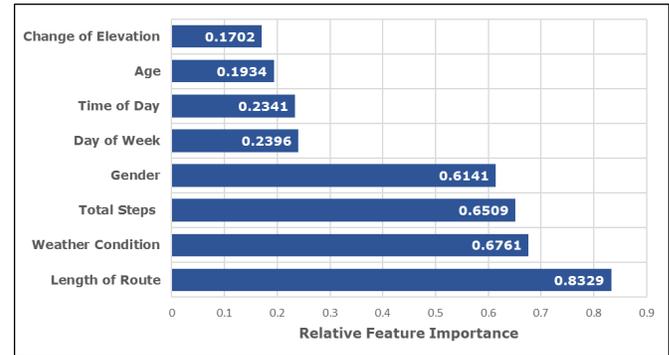
Figure 8. Relative Feature Importance in Random Forest method (for all features).

So far this research showed that with data collection (user profiling) and using predictive modeling techniques it is feasible to provide higher accuracy estimation of travel time for pedestrians. As it shown in figure 7 about 73% of actual correction values can be explained by Random Forest model. Almost all the features in the dataset can be automatically recorded or obtained from the smartphone. However, for myriad reasons accessing to some of the features might not be possible directly. For example, personal information such as age or gender can be obtained using Person class in Android SDK. However, the Person instances have values for age and gender only if the smartphone owner had set up the Gmail or Google+ accounts and filled all necessary information correctly (which is not the case most of the time). There are other ways for accessing personal data. Especially with the popularity of social networking apps and unified model of authentication, it is feasible to access the personal information via APIs provided by the major social networking apps (for example using Facebook's Graph API). However, using this approach needs the owner of a smartphone to have an account in the social network and has logged in his/her account on her device. There are some other ways to predict (estimate) the owner's age and gender by using machine learning methods. Recently researchers analyzed the mobile app choices of thousands of Android users to determine the predictability of certain attributes and found that installed apps and usage patterns can provide highly accurate insight on the user's gender, age, marriage status and even income [12]. However, the methods mentioned in that research are based on the assumption of installing of certain mobile apps. Nonetheless, the age and gender features are of major concern from privacy point of view.

4.3 Analytics with Less Private Data

As it was mentioned in previous section, gender and to some extent, age are important features for prediction of the correction value. As a result, we excluded age and gender from dataset and used the same machine learning techniques to predict the correction value. The results show that the prediction accuracy of all methods is reduced. On average there is 3.5% decrease in

prediction accuracy and the prediction accuracy of the Random Forest dropped by 5% (figure 9). This indicates the reasonable predictability strength of age and gender (especially gender) on the prediction of the correction value. The importance order of features (excluding age and gender) is similar to the figure 8. In comparison with figure 8, importance of length of the route, elevation change and weather condition are increased. This means in absence of the private features (gender and age) the predictive modeling method depends more on the other important features in order to keep predicting with high accuracy. This is also the reason for increasing the importance value of elevation change.
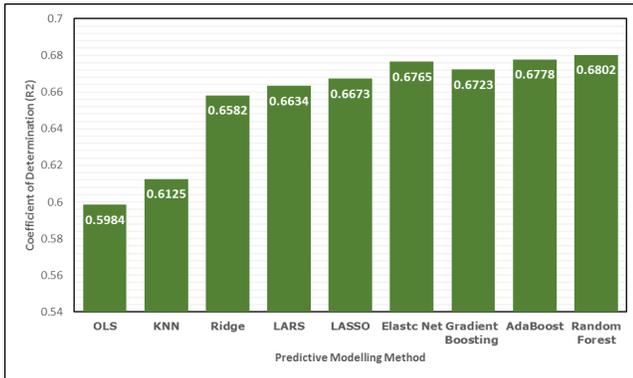

Figure 9. Prediction accuracy of various machine learning methods (all features except age and gender).

5. Conclusion

In this research, we evaluated the pedestrian mode of navigation apps in iOS, Android, and Windows Phone platforms. Through an experiment in Oxford, the UK we explored two major issues of Maps apps. Lack of information about location and type of pedestrian crossing is the first issue. In this context, we explored that neither of Maps app in iOS, Android and Windows Phone nor online mapping services (from Google and Microsoft) consider walkability of pedestrian routes in their routing algorithms (at least for case study area of this research). This issue is related to data quality and can be solved in various ways; from using crowd-sourcing data like OpenStreetMap to utilizing artificial intelligence for automatic detection of type and location of pedestrian crossing using street view services. Estimation of travel time for pedestrians is an important aspect of navigation services which in current form can cause several hours of delays in multi-modal travels. Learning from movement profile data of users is an opportunity for providing more accurate services for pedestrians. Unfortunately, neither major Maps apps provide such a personalized service (this is the second issue). In other words, Maps apps on smartphones are not smart enough to learn about the movement profile of their owners to provide higher accuracy services. As we illustrated in this paper, with predictive analytics it is possible to learn from movement profile, temporal and environmental data to provide more accurate and personalized estimation of travel time in pedestrian mode. With a real-world experiment we illustrated that the use of various machine learning algorithms and even without sensitive personal data, can lead to significant improvements in estimation of travel time.

6. REFERENCES

[1]   Amirian, P. and Basiri, A. 2016. Landmark-Based Pedestrian Navigation Using Augmented Reality and Machine Learning. *Progress in Cartography*. Springer International Publishing. 451–465.

[2]   Aylett, M.P. and Lawson, S. 2016. The Smartphone: A Lacanian Stain, A Tech Killer, and an Embodiment of Radical Individualism. *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems - CHI EA '16* (New York, New York, USA, 2016), 501–511.

[3]   Basiri, A., Amirian, P., Winstanley, A., Marsh, S., Moore, T. and Gales, G. 2016. Seamless Pedestrian Positioning and Navigation Using Landmarks. *Journal of Navigation*. (2016), 1–17.

[4]   Chang, K.-P., Wei, L.-Y., Yeh, M.-Y. and Peng, W.-C. 2011. Discovering personalized routes from trajectories. *SIGSPATIAL '11:Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Location-Based Social Networks*. 1, (2011), 33–40.

[5]   Chung, J. and Schmandt, C. 2009. Going my way: a user-aware route planner. *Proceedings of the 27th international conference on Human factors in computing systems - CHI 09* (New York, New York, USA, 2009), 1899.

[6]   Delling, D., Goldberg, A. V., Goldszmidt, M., Krumm, J., Talwar, K. and Werneck, R.F. 2015. Navigation made personal. *Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems - GIS '15*. (2015), 1–9.

[7]   Finnis, K.K. and Walton, D. 2008. Field observations to determine the influence of population size, location and individual factors on pedestrian walking speeds. *Ergonomics*. 51, 6 (Jun. 2008), 827–842.

[8]   Furukawa, H. 2015. Empirical evaluation of the pedestrian navigation method for easy wayfinding. *2015 International Conference and Workshop on Computing and Communication (IEMCON)* (Oct. 2015), 1–7.

[9]   Hastie, T., Tibshirani, R. and Friedman, J. 2009. *The Elements of Statistical Learning*. Springer Series in Statistics.

[10]   Kelly, C.E., Tight, M.R., Hodgson, F.C. and Page, M.W. 2011. A comparison of three methods for assessing the walkability of the pedestrian environment. *Journal of Transport Geography*. 19, 6 (2011), 1500–1508.

[11]   Letchner, J., Krumm, J. and Horvitz, E. 2006. Trip router with individualized preferences (trip): Incorporating personalization into route planning. *Proceedings of the National Conference on Artificial Intelligence*. 21, 2 (2006), 1795.

[12]   Malmi, E. and Weber, I. 2016. You Are What Apps You Use: Demographic Prediction Based on User's Apps. *arXiv preprint arXiv:1603.00059*. (2016).

[13]   de Montjoye, Y.-A., Hidalgo, C.A., Verleysen, M. and Blondel, V.D. 2013. Unique in the Crowd: The privacy bounds of human mobility. *Scientific reports*. 3, (2013), 1376.

[14]   Shahabi, C., Fan, L., Nocera, L., Xiong, L. and Li, M. 2015. Privacy-preserving inference of social relationships from location data. *Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems - GIS '15* (New York, New York, USA, 2015), 1–4.