Prediction of Landmarks Using (Personalised) Decision Trees

Eva Nuhn*, Sabine Timpf*

* Geoinformatics Group, University of Augsburg, Alter Postweg 118, 86159 Augsburg, {eva.nuhn, sabine.timpf}@geo.uni-augsburg.de

Abstract. Numerous studies claim that personal dimensions - such as *personal interests* or *prior spatial knowledge* - influence landmark selections for wayfinding. Nevertheless, up until now, a computational landmark salience model that includes personal dimensions has not been published. Thus, there has been no comparison possible between a conventional and a personalised model. In this paper, we provide such a comparison: We train two decision tree models - one personalised decision tree model (PdTm) and one conventional (CdTm) without personal information - to determine any differences between these models. We use the trees to predict selections of landmarks of participants in a case study. We evaluate the results and show that although the PdTm reacts sensitively to the personal dimensions it does not predict more landmarks than the CdTm.

Keywords. Landmarks, Decision Trees, Personalisation, Prior Spatial Knowledge, Personal Interests

1. Introduction

Our spatial memory is full of personal landmarks such as my working place or my doctor (Richter and Winter, 2014) or even brightly coloured doors, if it is our own (Lynch, 1960). Humans intuitively use landmarks with personal meaning especially in familiar environments (Sorrows and Hirtle, 1999). While human beings are able to easily provide such personalised landmarks it is much harder to get a routing application to do so. The data collection effort for the provision of personalised landmarks via an application is high and it raises the question if it is justifiable. To find an answer to this question we investigate the hypothesis: A model considering personal

LBS 2019
15th Conference on Location Based Services
Vienna, 11–13 November 2019

Published in "Adjunct Proceedings of the 15th International Conference on Location Based Services (LBS 2019)", edited by Georg Gartner and Haosheng Huang, LBS 2019, 11–13 November 2019, Vienna, Austria.

This contribution underwent double-blind peer review based on the paper. https://doi.org/10.34726/lbs2019.16 |

dimensions is able to correctly predict landmark selections more often than a conventional model without personal dimensions.

There are already a number of studies investigating which dimensions should be considered in such personalised landmark models (Nuhn and Timpf, 2017a, Nuhn and Timpf 2017b). In addition, there are studies showing that decision trees yield good results for landmark identification (Elias, 2006). In this paper we use a personalised decision tree model (PdTm) and a conventional one (CdTm) using the dimensions proposed in Nuhn and Timpf (2017b) to predict the selection of landmarks for route descriptions. We compare the results of a case study applying these models to test the hypothesis that a personalised decision tree model predicts significantly more landmarks than a conventional model. The CdTm is based on so called landmark dimensions (visual, semantic, and structural salience of objects (Sorrows and Hirtle, 1999)). A potential landmark might be salient because of outstanding visual attributes (e.g. colour or height). Visual salience is highly dependent on the surrounding objects. For example a yellow post box in a grey environment is highly salient. An object is semantically salient if it has an outstanding meaning. It might have cultural or historical importance or show explicit marks (Raubal and Winter, 2002). Highly accessible objects with a prominent location (e.g. squares) are structural salient. The PdTm includes, in addition to landmark dimensions, also personal dimensions. There are several personal dimensions influencing landmark salience (Nuhn and Timpf, 2017b). Amongst them: *prior spatial knowledge* and *personal interests* (Nuhn and Timpf, 2017a). Several studies confirm the importance of spatial knowledge for landmark predictions (Hamburger and Röser, 2014, Quesnot and Roche, 2015). Inspired by Siegel and White (1975), Nuhn and Timpf (2017b) introduced four attributes to consider prior spatial knowledge of a traveller: *no knowledge*, *landmark knowledge*, *route knowledge,* and *survey knowledge*. The second important dimension is personal interests, which guides attention and, thus, results in the perception of objects and configurations (Rensink et al., 1997). Personal interests reflect person-specific orientation and provide important categories for action goals in a situation where persons are free to do as they please (Krapp et al., 2014).

In this paper, we first describe the data collection and preparation process for the computational models. Subsequently follows the description of the training of PdTm and CdTm, including the identification of optimised model parameters. Afterwards, the results of the trees on case study data are compared and discussed. The paper closes with conclusions and an outlook on future work.

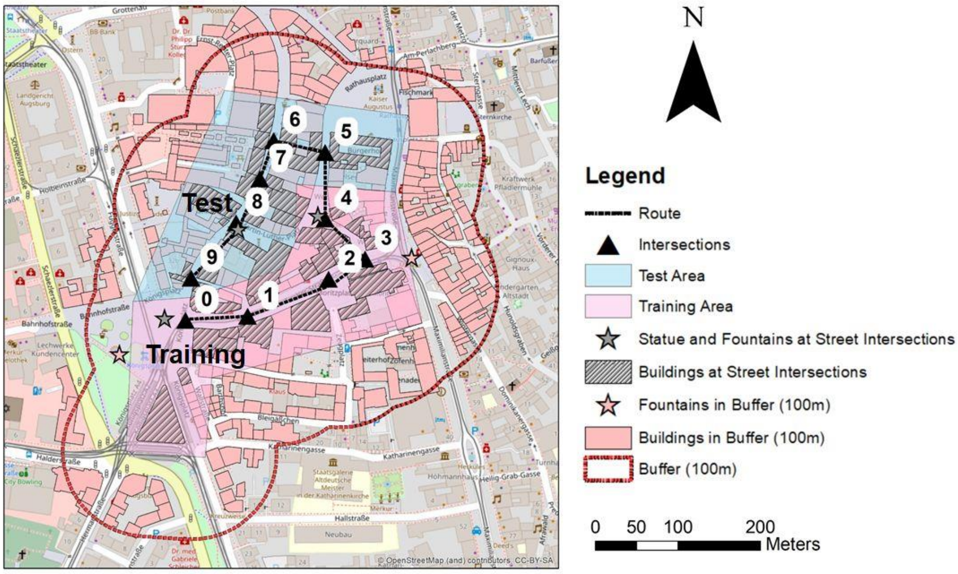

Figure 1. Intersections along the route.

2. Data Collection and Preparation – Case Study

For data collection we chose to concentrate on a route in the inner city of Augsburg because it includes different objects belonging to different topics of personal interests. The route is around 640 meters long and includes 10 street intersections (Figure 1). The objects at the street intersections are 44 buildings, two fountains, and a statue. The modelling of the landmarks for consideration in the trees requires a number of data sources. Additionally to information on the landmark itself we need information on the personal dimensions.

2.1. Landmark Dimensions

Landmark dimensions include information on visual, semantic, and structural dimensions of the objects at the street intersections (Nuhn and Timpf, 2018). In this study we use data from OSM (OpenStreetMap) and an official 3D city model. Further data (mainly visual data, such as colour) was collected during a field study. Objects have salience if they are different from the surrounding objects (e.g. in a 100m buffer as proposed by Raubal and Winter (2002) (Figure 1)). We calculate the salience for the landmark dimensions following the formulas provided in Nuhn and Timpf (2017a, 2018).

2.2. Personal Dimensions

We collected personal dimensions, interest and spatial knowledge, as well as information on landmarks in the framework of a case study. Decision trees need objects for training, which classify as a *landmark* and so called *NALs* (an object which does not classify as a landmark). 51 people participated in the case study, 24 of whom are female. The mean age of the participants is 33.1 years (min = 19, max = 73). 23 participants live in Augsburg, seven of them since their early childhood (age ≤10) or birth, and, thus, are spatially familiar. Six participants are not born in Germany. We use ESRIs Survey123 for data collection. The app allows to create and publish survey forms (Survey123, 2018). For the study, we set up a survey with questions about personal interests, prior spatial knowledge, and about the objects at the street intersections along the route. Participants rate their interest in shopping, culture, historical monuments, and gastronomy on a Likert scale with *no = 1*, *low = 2*, *medium = 3*, *high = 4*, and *very high = 5* items. Participants walked along the route and stated at each street intersection if they have been there before or not. In case they answered affirmative, they are asked about their spatial knowledge in the area of the intersection (landmark, route, or survey knowledge). In case they have never been at the street intersections, the questions include an additional question about no knowledge (Table 1). Afterwards, participants were asked to do their object selections. Survey123 provides photos of the objects at the street intersections. The photos are only intended as an aid for identifying objects in the real environment. Participants are encouraged to look at the real objects to do their selections. Because we assume that direction givers adapt their directions to the expected personal interests and spatial knowledge of the recipient, not to their own preferences, we told participants that they should imagine personally addressed route directions. Based on this assumption they had to select an object they like (landmark) and one object they don't like (NAL) for such a route direction. In total, 47 objects are presented with a mean of 4.7 (min = 4, max = 6) objects per street intersection.

SPspK	Street intersection	Area	ø
1	Yes	Survey	21.9
2		Route	12.1
3		Landmark	8.2
4	No	Survey	0
5		Route	0
6		Landmark	1
7		No	7.8

Table 1. Stages of spatial knowledge and average number of selection.

The result of the study is a corpus of landmarks and NALs. The Survey123-App presented the same objects for landmarks and NALs, which resulted in some cases in the same object being selected for both instructions. For further analysis, only those street intersections were kept where two different objects had been selected for both (landmarks and NALs). This resulted in 503 *landmarks* and the same number of *NALs*. Ratings for topics of interest and information about spatial knowledge for the street intersections are available for all participants.

3. Decision Tree Training

We use the decision tree algorithm CART (Classification and Regression Trees) (Breiman et al., 1984) in this work. CART might grow until it perfectly classifies a data set. However, this may lead to overfitting. In this case the tree tightly fits the data set so well that is is inaccurate in predicting the outcomes of previously unseen data. Decision trees are almost always stopped before they are fully grown to avoid overfitting. There are various parameters that help to decide when to stop growing (Scikit, 2018). The *criterion* measures the quality of the split (available functions are the GINI index (Breiman et al., 1984) or entropy (Quinlan, 1986)). *Splitter* is a method to split the node, it is divided into 'best' or 'random'. The *minSamplesSplit* is the minimum number of samples required to split a tree node, whereas *minSamplesLeaf* is the minimum number of samples required to be at a leaf. Finally, *maxDepth* determines the maximum depth of the tree. There are also other training parameters considering weights for data entries or target variables (*landmark* or *NAL*). We decided not to introduce weights and to restrict ourselves to the five parameters described here.

The input dataset for the trees includes landmark and NALs with values for landmark and personal dimensions. The CdTm considers only the landmark dimensions (visual, semantic, and structural), whereas PdTm considers landmark as well as personal dimensions. The available data is used to train both decision tree models as well as to test them. We divide our data set consisting of data for the 10 street intersections into two sets of equal size: training and test area (Figure 1). The training set includes 252 landmarks and 252 NALs. There are combinations of spatial knowledge and personal interests ratings from the training set not appearing in the test set. In order not to influence the prediction we excluded landmarks with these combinations from the test set. This results in a test set with 232 landmarks. We do not consider NALs for testing, because here we are only interested in landmark prediction.

Parameter	Coarse	PdTm	CdTm	Finer	PdTm	CdTm
Criterion	Gini,Entropy	Entropy	Gini	Gini,Entropy	Gini	Gini
Splitter	Best,Random	Random	Best	Best,Random	Random	Random
minSamplesSplit	[5,10,…,50]	30	5	PdTm: [25,26,..,35] CdTm: [2,3,,10]	34	2
minSamplesLeaf	[5,10,…,50]	5	5	[1,2,...,10]	5	1
maxDepth	[5,10,…,50]	10	5	PdTm: [5,6,..,15] CdTm: [1,2,..,10]	9	4
Average Accuracy [%]		76.78	76.19		77.38	76.19

Table 2. Parameter values for initial coarse grid-search (middle) and for finer grid-search (right).

However, the number of data items for training might be too small to gain reliable results. A solution for this problem is cross-validation (Stone, 1974). We use stratified cross-validation, which divides the data set in disjoint classes with equal class distributions (Kohavi, 1995). According to Borra and Di Ciaccio (2010) a reliable result can be obtained with k=10. A widely used method to identify optimal parameter values for the parameters defined above combines cross-validation with grid-search (Chicco, 2017). We implement the (P/C)dTm as a Toolbox in ESRI's ArcGIS 10.5.1 using Python 2.7.12. In addition, we use statistic packages to train and test the trees (Pedregosa et al., 2011). The packages provide methods for grid-search and cross-validation. We start with a coarse grid-search with 10-fold stratified cross-validation to train the trees. Table 2 shows the initial parameter settings. The coarse grid-search identifies the highest average accuracy for the PdTm for the values in Table 2 with a score of 76.78. The average accuracy for the CdTm is with a value of 76.19 slightly lower. Next, we conduct a finer grid-search, varying the parameters of *minSamplesSplit*, *minSamplesLeaf,* and *maxDepth* around the best values (see Table 2). The accuracy of PdTm improves and reaches a value of 77.38, whereas the accuracy of the CdTm stays the same (76.19). After identifying the best parameters, we build the final decision trees on the training set. Figure 2 and 3 show the resulting trees.

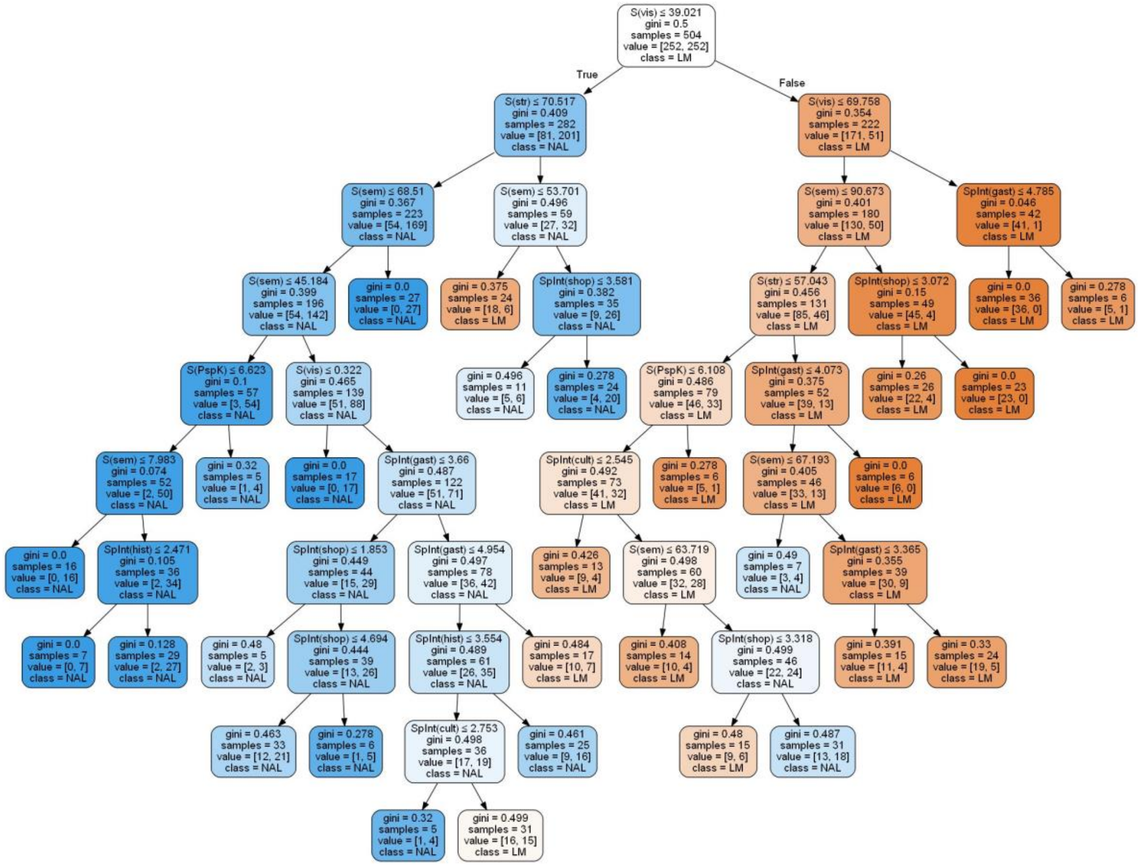

Figure 2. Trained PdTm.

4. Comparison and Discussion of the Results

This section compares the landmarks determined by the trees with landmarks selected by the study participants. We apply our final model with the parameters in Table 2 to the test set. Remember, that we do not consider NALs for testing, because here we are only interested in landmark prediction. A performance measure considering only landmarks is the recall (Buckland and Gey, 1994). The CdTm identifies 157 landmarks of the test set, or a recall of 67.67%, and the PdTm identifies 154 landmarks, or 66.38% recall on the test set. Thus, the CdTm identifies slightly more landmarks than the PdTm. We investigate these findings with a subsequent McNemar's test (McNemar, 1947) to find out whether this difference is significant. The McNemar's test analyses the results of a study where two different models are applied to the same objects. The test operates upon a contingency table, which relies on the fact that both trees are trained on exactly the same training set and evaluated on exactly the same test set (Brownlee, 2018).

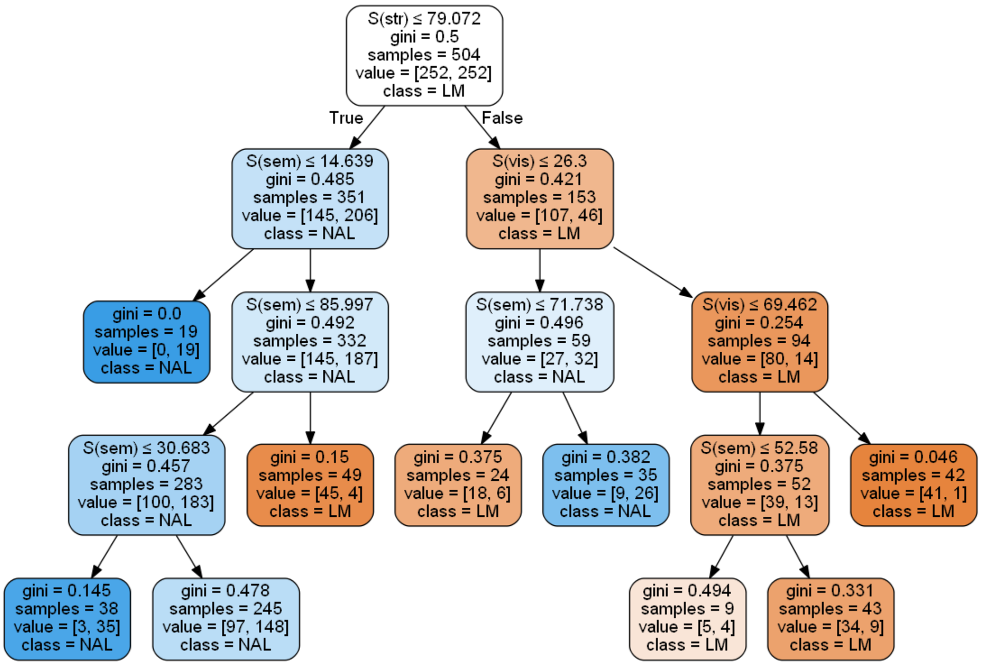

Figure 3. Trained CdTm.

The null hypothesis of McNemar's test claims that the two models have the same error rate (H_0: CdTm = PdTm). In case the null hypothesis is rejected, it suggests that there is evidence to suggest that the two models have different performance (H_A: CdTm ≠PdTm) (Dietterich, 1998). The two-tailed p-value equals 0.810 with a McNemar's test statistic of 0.058. By conventional criteria (significance level of 5%), this difference is considered not to be statistically significant. This means, the CdTm does not identify significantly more landmarks than the personalised model.

The CdTm considers all landmark dimensions (visual, semantic, and structural). The structural dimension appears on the first level in the tree, followed by the semantic and the visual dimension (second level). The tree is sensitive to a change in the input data of semantic, visual, and structural salience (not evaluated further at this point). The ability of CART to use the same dimensions more than once in different parts of the tree is reflected in the CdTm. For example s_{sem} appears in different branches of the tree. First a coarse division of the data in $s_{str} \leq 79.072$ is done on level 1 and a finer division in $s_{sem} \leq 14.639$ follows on level 2. Visual salience also appears more than once. Figure 3 shows, that the tree generates terminal nodes with the same class (e.g. left *NAL*). The algorithm does not stop earlier because *minSamplesSplit*, *minSamplesLeaf*, or *maxDepth* is not reached. Because we determined minSamplesSplit = 2, minSamplesLeaf = 1, and maxDepth = 4 (compare Table 2) the algorithm stops before it can yield all pure leaf nodes. In Figure 3 the terminal node on the left shows 3 samples of the class *Landmark* and 35 samples of class *NAL*. In case decision tree growing would stop already after splitting in Landmark and *NAL* (level 4 in Figure 3) it would produce a terminal node with 100 samples belonging to class *Landmark* and 183 objects belonging to the class *NAL*, which would be far less useful. The terminal node on the right is less pure than the terminal node on the left. It shows 97 samples of the class landmark and 148 samples of the class NAL. Thus, a number of objects which are actually selected as landmarks by the study participants end up in this node and are therefore predicted as NALs. However, the finer grid-search identifies the model parameters of the CdTm in Table 2 as the ones yielding the highest average accuracy. Consequently, we use the CdTm trained on these parameters in this work.

The root node of the PdTm starts with analysing $s_{vis} \leq 39.021$. This fundamental division is followed by s_{str} and s_{vis} on the second level. The PdTm shows sensitivity to the inputs of the landmark dimensions as well as to the inputs of the personal dimensions (not evaluated further at this point). The input values of these dimensions decide if an object becomes a *landmark* or a *NAL*. CART uses also the same dimensions more than once in different parts of PdTm (compare Figure 2). However, most of them appear with the same decision. Only $s_{pInt(cult)}$ behaves differently. It appears twice with similar thresholds but contradictory decisions. Nevertheless, this is comprehensible because whether the PdTm predicts an object as a landmark is also dependent on the values of the other dimensions. The PdTm shows also branches, which generates leaves with the same class, due to the fact that the tree size is dependent on the parameters obtained with the finer grid-search (Table 2).

The PdTm makes a distinction between s_{PspK} = 7 and the other ratings. Table 1 shows the average number of selections of the spatial knowledge ratings at the street intersections. It reveals that study participants did not choose the ratings 4 and 5 at all. In addition, on average only one study participant chose s_{PspK} = 6 at a street intersection. This indicates that these ratings do not influence the splitting of the PdTm. Thus, the distinction between s_{PspK} = 7 (no familiarity at all) and all the other ratings (familiarity) seems to be plausible.

An interesting fact is that the PdTm splits for the personal interests $s_{pInt(cult)}$, $s_{pInt(shop)}$, and $s_{pInt(hist)}$ either between s_{pInt} = 2 and s_{pInt}= 3 or s_{pInt}= 3 and s_{pInt}= 4. This suggests that the tree identifies a difference between a study participant which is interested and which is not. We observe, that the *medium* rating is either assigned to the lower rating or to the higher ratings. This might be explained by *survey optimising* (Krosnick, 1991), which describes an behaviour occurring under cognitive load and when study participants attempt to be fully diligent. Consequently, they try to avoid this effort but they want to answer responsibly (Krosnick, 1991, Krosnick and Fabrigar, 1997). The result is that, the personal interests rating *medium* might be either chosen by a study participant who is actually interested in a topic as well as by a participant who is not. $s_{pInt(gast)}$, on the other hand, is an exception: it splits between *very high* and all the other ratings, suggesting that there is a difference in landmark selection between someone with a very high interest in gastronomy and all the others.

5. Conclusions

In this study we trained two decision trees - one with personal information and one without. We carried out k fold cross-validation with grid-search and determined optimal model parameters. Then, we built the final trees with these parameters and use them to predict selections of landmarks of the participants of a study. We evaluated the results and identified - contrary to our hypothesis - that there is no significant difference between a CdTm and a PdTm. According to these results, we have to reject our hypothesis.

There might be a variety of causes for this result. The most obvious interpretation is that personal dimensions are just not important for landmark selections. This would confirm the findings of Gramann et al. (2017) who also find that directions including information of personal interests associated with landmarks did not perform better than non-personalised directions including irrelevant information about landmarks. However, there might be a number of other reasons for this result:

- **Dimensions.** We considered landmark and personal dimensions. However, there might be missing landmark dimensions influencing the CdTm as well as other dimensions such as e.g. an environmental dimension (Nuhn and Timpf, 2017a).
- **Methods**. Other models, besides decision trees might be useful as well. This includes models inspired by theory (e.g. the model proposed by Raubal and Winter (2002)) as well as other machine learning models.
- **Overall model**. In this work we investigated one overall model to predict landmark selections. Another possible approach could be individual models for each study participant. As study participants might be influenced by individual intangible parameters resulting in individual landmark selections, which might not be covered by an overall approach.

This investigation of possible reasons for the rejection of the hypothesis reveals a number of open research questions for future work. However, we currently have to conclude that the data collection effort for obtaining information on spatial knowledge and personal interests for an applied system might not be justifiable. In case future work will confirm these findings it is most likely sufficient to focus on existing conventional landmark prediction models and to concentrate on their use in applied pedestrian wayfinding applications.

Acknowledgements

The author is grateful for financial support of the Young Researchers Travel Scholarship Program of the University of Augsburg.

References

Borra, S. and Di Ciaccio, A., 2010. Measuring the prediction error. a comparison of cross-validation, bootstrap and covariance penalty methods. Computational statistics & data analysis 54(12), pp. 2976–2989.

Breiman, L., Friedman, J., Stone, C. and Olshen, R., 1984. Classification and Regression Trees. Taylor & Francis.

Brownlee, J., 2018. Statistical methods - how to calculate mcnemars test to compare two machine learning classifiers. Accessed December 2018.

Buckland, M. & Gey, F. (1994), 'The relationship between recall and precision', Journal of the American society for information science 45(1), 12–19.

Chicco, D., 2017. Ten quick tips for machine learning in computational biology. BioData mining 10(1), pp. 35.

Dietterich, T. G., 1998. Approximate statistical tests for comparing supervised classification learning algorithms. Neural computation 10(7), pp. 1895–1923.

Elias, B., 2006. Extraktion von Landmarken f¨ur die Navigation. PhD thesis, University of Hanover, Institute of Cartography and Geoinformatics.

Gramann, K., Hoepner, P. and Karrer-Gauss, K., 2017. Modified navigation instructions for spatial navigation assistance systems lead to incidental spatial learning. Frontiers in psychology 8, pp. 193.

Hamburger, K. and Röser, F., 2014. The role of landmark modality and familiarity in human wayfinding. Swiss Journal of Psychology 73(4), pp. 205–213.

Kohavi, R., 1995. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: Int. Joint Conference on Artificial Intelligence, pp. 1137–1145

Krapp, A., Hidi, S. and Renninger, A., 2017. Interest, learning and developement. In: R. Ann, H. Suzanne and K. Andreas (eds), The role of interest in learning and development, Psychology Press, pp. 3–26.

Krosnick, J. A., 1991. Response strategies for coping with the cognitive demands of attitude measures in surveys. Applied Cognitive Psychology 5(3), pp. 213–236.

Krosnick, J. A. and Fabrigar, L. R., 1997. Designing rating scales for effective measurement in surveys. In: L. Lyberg, P. Biemer, M. Collins, E. de Leeuw, C. Dippo, N. Schwarz and D. Trewin (eds), Survey Measurement and Process Quality, John Wiley & Sons, Inc., pp. 141–164.

Lynch, K., 1960. The image of the city. MIT press.

McNemar, Q., 1947. Note on the sampling error of the difference between correlated proportions or percentages. Psychometrika 12(2), pp. 153–157.

Nuhn, E. and Timpf, S., 2017a. A multidimensional model for selecting personalised landmarks. Journal of Location Based Services 11(3-4), pp. 153–180.

Nuhn, E. and Timpf, S., 2017b. Personal dimensions of landmarks. In: A. Bregt, T. Sarjakoski, R. van Lammeren and F. Rip (eds), Societal Geo-innovation, Springer International Publishing, pp. 129–143.

Nuhn, E. and Timpf, S., 2018. An overall framework for personalised landmark selection. In: P. Kiefer, H. Huang, N. Van de Weghe and M. Raubal (eds), LBS 2018: 14th International Conference on Location Based Services, Springer, pp. 231–253.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A. & Cournapeau, D. (2011), 'Scikit-learn: Machine learning in python', Journal of Machine Learning Research 12, 2825–2830.

Quesnot, T. and Roche, S., 2015. Quantifying the significance of semantic landmarks in familiar and unfamiliar environments. In: S. I. Fabrikant, M. Raubal, B. Michela, C. Davies, S. Freundschuh and S. Bell

Quinlan, J. R., 1986. Induction of decision trees. Machine learning 1(1), pp. 81–106.

Raubal, M. and Winter, S., 2002. Enriching wayfinding instructions with local landmarks. In: M. J. Egenhofer and D. M. Mark (eds), GIScience 2002. LNCS vol 2478, Springer, Berlin, Heidelberg, pp. 243–259.

Rensink, R. A., O'Regan, J. K. and Clark, J. J., 1997. To see or not to see: The need for attention to perceive changes in scenes. Psychological science 8(5), pp. 368–373.

Richter, K.-F. and Winter, S., 2014. Landmarks – GIScience for Intelligent Services. Springer International Publishing.

Scikit, 2018. Documentation of skckit-learn 0.19.2. Accessed August 2018.

Siegel, A. W. and White, S. H., 1975. The development of spatial representations of large-scale environments. Advances in child development and behavior 10, pp. 9–55.

Sorrows, M. E. and Hirtle, S. C., 1999. The nature of landmarks for real and electronic spaces. In: C. Freksa and D. M. Mark (eds), COSIT 1999. LNCS vol 1661, Springer, Berlin, Heidelberg, pp. 37–50.

Stone, M., 1974. Cross-validatory choice and assessment of statistical predictions. Journal of the royal statistical society. Series B (Methodological) 36(2), pp. 111–147.

Survey123, 2018. Survey123 for arcgis. Accessed November 2018.