Deriving Salience Models from Human Route Directions

Jana Götze and Johan Boye

KTH, School of Computer Science and Communication, 100 44 Stockholm, Sweden

Abstract. We present an approach to derive individual preferences in the use of landmarks for route instructions in a city environment.[1] Each possible landmark that a person can refer to in a given situation is modelled as a feature vector, and the preference (or *salience*) associated with the landmark can be computed as a weighted sum of these features. The weight vector, representing the person's personal salience model, is automatically derived from the person's own route descriptions. Experiments show that the derived salience models can correctly predict the user's choice of landmark in 69% of the cases.

1 Introduction

Automatically providing real-time route instructions to city pedestrians is an increasingly important problem, as more and more people have smartphones with GPS receivers. Such wayfinding systems use data from a geographic database to construct a route from the user's starting position to his stated goal, and then give the instructions as the user is moving: When the user reaches a node p_i in the planned route, the system informs the user how he should go to get to the next node p_{i+1}. Obviously, it is vital that each instruction is unambiguous and understandable, lest the user takes a wrong turn.

It would be preferable if wayfinding systems would base their instructions on *landmarks*, by which we understand distinctive objects in the city environment, since it is well established that it is predominantly by landmarks people describe routes to one another (see e.g. [2]). However, even on this basic premise, there are a number of options to consider. At each decision point, there are a number of possible landmarks to choose from, and which one(s) to use in a specific route instruction is a difficult problem. In the literature, it is generally assumed that the candidate landmarks can be assigned a *salience* measure, by which they can be compared, and the most salient features are also the most suitable to use in route descriptions. Many researchers have proposed schemes for computing salience from a variety of factors (see e.g. [3, 6, 9]).

In this article, we investigate to what extent salience computations can be data-driven, that is, (semi-)automatically estimated from human route descriptions. Our aim is to create empirically motivated *personalized* salience models, and integrate them into our spoken-dialogue system for city exploration [1]. Two

[1] Supported by the European Commission, project *Spacebook*, grant no 270019.

hypotheses underlie our work: Firstly, that salience is *user-dependent*. Secondly, if a user is asked to give a routing instruction in a specific situation, he would do so using the landmarks he himself thinks are most salient.

The second hypothesis suggests a kind of tuning mechanism for a wayfinding system: Before being guided by the system, the user first walks around and describes the way he is going by means of landmarks. The system interprets the user's descriptions and uses them to derive a personalized salience model, which can later be used when guiding the same user in other parts of the city. The present paper presents a preliminary study showing that this idea is indeed viable.

2 Deriving Salience Models

For the learning of salience models, we use the Large Margin Algorithm, introduced in [4]. Each landmark can be described as a vector of numerical features, $\mathbf{x} = (x_1, \ldots, x_n)$ specifying costs along n dimensions. The dimensions might represent scalar attributes such as distance, or categorical attributes (e.g. 1 if the landmark is a restaurant, 0 if it is not). The salience $s(\mathbf{x})$ is a linear combination $\mathbf{w} \cdot \mathbf{x}$, where $\mathbf{w} = (w_1, \ldots, w_n)$ is the salience model that specifies the relative importance of the different features for the user. Naturally we do not assume that the user knows the values of his salience model, or indeed even that such a model exists. Instead we automatically infer the model as follows:

Whenever a person uses a landmark A in a description, he is preferring A over a number of other candidates that *could have been* used in the description but were not. That is to say that A has a lower cost according to the person's personal salience model than has any other candidate B, i.e. $\mathbf{w} \cdot (\mathbf{x_B} - \mathbf{x_A}) > 0$, where $\mathbf{x_A}$ and $\mathbf{x_B}$ are the vectors representing A, and B, respectively. Each route description from the user involving a landmark thus generates a number of inequalities, all in the form $\mathbf{w} \cdot (\mathbf{x_{B_i}} - \mathbf{x_{A_i}}) > 0$, for $1 \leq i \leq m$. Our goal is to find appropriate values for the weights in $\mathbf{w}$ that satisfy all these inequalities. This can be done by solving the following linear optimization problem, e.g. with the Simplex method [7]:

$$\begin{array}{lll} \text{minimize} & \sum_{j=1}^{n} w_j & \\ \text{subject to} & \mathbf{w} \cdot (\mathbf{x_{B_i}} - \mathbf{x_{A_i}}) \geq 1, & 1 \leq i \leq m \\ & w_j \geq 0, & 1 \leq j \leq n \end{array}$$

This formulation of the problem assumes that a person is always consistent in his preferences. For the case he is not, we use a slightly extended version of the basic Large Margin Algorithm (see [4] for details).

3 Problem Description and Encoding

Consider the example in Figure 3. The figure shows a situation in one of our experiments where the subject chooses to describe the way using a supermarket,

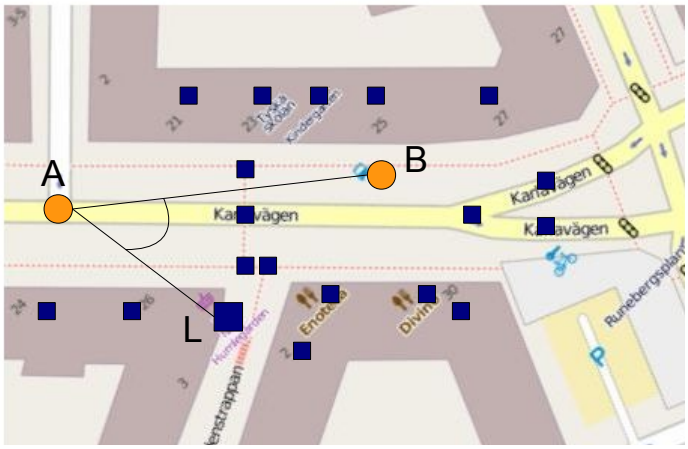

Fig. 1. An example route segment from A to B. The squares represent the landmarks in the contexts of A and B. L represents a landmark referred to by the user (a supermarket).

indicated by the larger square: *"and then when you've reached a crossroad [. . .] you turn to your left and you'll see there's gonna be an* ICA*, a foodstore, and a little bit further down the road there's gonna be a bus stop"*. In the figure, the "crossroad" is indicated by "A", and the bus stop by "B".

Every landmark belongs to the *context* of its closest node. When describing the way from A (the *starting node* of the segment) to B (the *goal* node of the segment), all landmarks in the contexts of these two nodes are possible referents. We will refer to this set of landmarks as the *candidate set* for A and B. This set is visualized as square-shaped icons in the figures. The candidate set is obtained from the OpenStreetMap (OSM) geographic database [5].

The method described in Section 2 requires every landmark L to which the user can refer to be modelled as a vector of features. In this experiment, we use a vector of 12 features that are computable from our geographic database. These features form an initial set of structural landmark features [8] and we are planning to further explore which other features are important for computing salience. The features used here are the following:

- **Distance** between the user's position A and the landmark L.
- **Distance** between the landmark L and the goal node B.
- **Angle** between the lines AL and AB.
- **Name**: Categorial attribute having the value 1 if the landmark has a name (e.g. "7-Eleven"), or belongs to something that has a name, e.g. a node on X street, and 0 otherwise.
- **Type**: These 8 features represent the type of the landmark according to whether they belong into the categories *road network*, i.e. the landmark node is part of a street, *building*, *eating & pleasure*, e.g. a restaurant or a theater,

shops, *entrances*, i.e. a specific house number on a street, *areas*, e.g. a park or a construction site, *structures*, e.g. a statue or a fountain, or *other*. Each landmark is of at least one type, which is indicated by the value 1 in the corresponding slot.

In the example in Figure 3, the supermarket that is referenced by the user (the larger square), is represented by the vector $(5.0, 5.0, 40, 1, 0, 0, 0, 1, 0, 0, 0, 0)$. The first two positions contain the distances (the 2-logarithm of the actual distance in metres, rounded to the nearest integer). The third position represents the angle (in degrees). The '1' in the succeeding slot indicates that the landmark has a name "ICA". The values in the final 8 slots indicate that the landmark is a shop, but no other type.

4 Data Collection

A number of subjects (engineering students) were asked to describe a route to someone unfamiliar with the area, imagining that they were talking to this person on the phone. The subjects had just walked the same route themselves and should therefore remember it well. To further help them recall their trajectory, they were also shown their route on a map on the screen by a moving mouse cursor (i.e. without using speech), and they could also look at the map while they described the route.

The subjects' speech was recorded and segmented according to route segments before transcription. Each route segment starts at a node A and ends at a goal node B. The nodes A and B were inferred from the subjects' instructions, as they used phrases like *"and when you are at the intersection, turn left and walk until the bus stop"*. While the route as a whole differed only slightly from subject to subject, the routes do not necessarily consist of the same number of segments. The segmentation here is derived from the subjects' descriptions. Each segment was also annotated with all landmarks in the database that the subject referred to. The set of landmarks used by the subjects often includes the goal node B itself, as in the example in Figure 3. In that example, the instruction was annotated with the node representing the supermarket and the node representing the bus stop. It can also be the case that the goal node B is not mentioned explicitly, as in *"and when you are at the traffic light, cross street S"*. In this case, the goal node B is implicit, and not part of the landmarks referred to by the subject.

Prior to describing the route, the subjects had walked them themselves, following instructions given by our prototype system. This means that their own instructions might be influenced by what they just heard. However, the system's instructions only partly used landmarks and otherwise relied on relative instructions such as "turn left". This strategy sometimes resulted in ambiguous or wrong instructions, and the subjects were asked to "improve upon the system's behavior".

For each subject, we thus have a number of annotated segments, each consisting of a start node, an end node, and at least one landmark that the subject

referred to (his *preferred* landmark(s) in this segment). Segments where the subject didn't refer to anything at all were excluded from this experiment. The candidate set for the segment (i.e. the landmarks the user *could have* referred to) was automatically computed from the OSM database and contains on average 22 landmarks.

The preferred landmarks might or might not be part of the candidate set. There are two possible reasons for a preferred landmark not to be part of the candidate set: Either the user referred to something that is not in the database at all (in which case we removed the reference), or he referred to something that is farther away, and doesn't belong to the context of neither A nor B (this latter case actually never happened in our experiments).

An *instance*, of the salience model learning problem, then, is a candidate set together with one or several preferred landmarks, at least one of which is part of the candidate set. The set of all instances for a particular user was split into a training set and a test set. The training set was used to derive a salience model $\mathbf{w}$ according to the method presented in Section 2. To evaluate $\mathbf{w}$, the salience of each member of each instance of the test set was computed. A *successful* instance is one in which one of the preferred landmarks had the best salience according to $\mathbf{w}$. The number of successful instances in the test set is an indicator of how well the learned salience model actually reflects the preferences of the user.

5 Results

The results are presented in Table 1. For all individual salience models, at least half of the test instances are successful. In one case, the model even returns all the instances as successful. To get an insight into how well the models perform on those landmarks that did not receive the lowest cost but were used by the subject, we also compute the measure RANK. For this measure, we compute the percentage of landmarks receiving costs that were equal or higher than the preferred landmark's cost (recall that the lower the cost, the more salient the landmark). The number of landmarks that can be referred to differs depending on the particular route segment and this measure reflects how high the salience model ranked a landmark in comparison to all available landmarks. For example, subject 1's model has two successful test instances, and in the other two ranks the preferred landmark as 3 of 14 in one instance, and as 5 of 39 in the other.

6 Discussion

The results are encouraging insofar that in 69%, the method actually managed to mimic the user's own salience preferences, although the model is built from very few training examples. Note that the ratio of training vs. testing segments differs between the subjects. Initially, the training set contains two thirds of the route segments. For some subjects, the training size had to be reduced, because our algorithm is limited in the number and size of route segments it can process.

Table 1. For evaluation, we used the induced weights to compute costs on test sets and counted in how many cases the best option was a landmark used by the subject, including also reference to streets and squares. SEGMENTS: total number of route segments, TESTS: number of test instances, SUCC: number (and percentage) of successful test instances, RANK: percentage of landmarks with equal or higher cost

SUBJ	SEGMENTS	TESTS	SUCC	RANK
1	13	4	2 (0.50)	0.93
2	16	5	3 (0.60)	0.94
3	9	3	2 (0.67)	0.94
4	9	3	2 (0.67)	0.94
5	16	10	7 (0.70)	0.95
6	12	4	4 (1.00)	1.00
total	75	29	**20 (0.69)**	0.95

Future work includes a user study in which users are recorded as they walk around the city describing their environment in real-time (rather than describing a route after having walked it). We also plan to analyse in detail whether the individual preference models all have something in common (i.e. whether there are general properties of salience models that always hold). The results of such an analysis might allow us to restrict our candidate sets, thereby making it possible to build the models from more examples. Furthermore, we aim to investigate which other features, apart from the ones we are considering in this article, are important for the salience computation problem.

References

1. Boye, J., Fredriksson, M., Götze, J., Gustafson, J. and Königsmann, J. (2012) Walk this way: Spatial grounding for city exploration. *Proc. 4th IWSDS*
2. Denis, M., Pazzaglia, F., Cornoldi, C. and Bertolo, L. (1999) Spatial discourse and navigation: an analysis of route directions in the city of Venice. *Applied cognitive psychology*, vol 13, no 2.
3. Duckham, M., Winter, S. and Robinson, M. (2010) Including landmarks in routing instructions. *Journal of Location Based Services*, vol. 4, no. 1, pp. 28–52.
4. Fiechter, C-N. and Rogers, S. (2000) Learning subjective functions with large margins. *Proc. 17th ICML*, pp. 287–294.
5. Haklay, M. (2008) OpenStreetMap: User-generated street maps. *Pervasive computing IEEE*, vol. 7, issue 4, pp. 12–18.
6. Nothegger, C., Winter, S. and Raubal, M. (2004) Selection of salient features for route directions. *Spatial cognition and computation*, 4(2), pp. 113–136.
7. Papadimitrou, C. and Steiglitz, K. (1982) *Combinatorial optimization: Algorithms and complexity*, Prentice-Hall.
8. Sorrows, M.E. and Hirtle, S.C. (1999). The nature of landmarks for real and electronic spaces. *Spatial information theory: Cognitive and computational foundations of geographic information science*, vol. 1661 LNCS, pp. 37–50.
9. Xia, J., Richter, K-F., Winter, S. and Arnold, L. (2011) A survey to understand the role of landmarks for GPS navigation. *Proc. PATREC research forum.*