Learning Landmark Salience Models from Users' Route Instructions

Jana Götze & Johan Boye

KTH Royal Institute of Technology
School of Computer Science and Communication
10044 Stockholm, Sweden
{jagoetze,jboye}@kth.se

Abstract. Route instructions for pedestrians are usually better understood if they include references to landmarks, and moreover, these landmarks should be as salient as possible. In this paper, we present an approach for automatically deriving a mathematical model of salience directly from route instructions given by humans. Each possible landmark that a person can refer to in a given situation is modeled as a feature vector, and the salience associated with each landmark can be computed as a weighted sum of these features. We use a ranking SVM method to derive the weights from route instructions given by humans as they are walking the route. The weight vector, representing the person's personal salience model, determines which landmark(s) are most appropriate to refer to in new situations.

1. Introduction

Recently there has been an increasing interest in systems capable of providing natural language route instructions to pedestrians in a city environment (Rehrl et al., 2010; Janarthanam et al., 2012; Google, 2013; Boye et al., 2014). Such systems track the pedestrian's position using the GPS on his smartphone, and can therefore produce real-time instructions like "turn left here" or "now you should walk towards the cafe on the corner". Obviously, a recurring challenge for such wayfinding systems is to find the best formulation of the next instruction, minimizing the risk of a misunderstanding.

When giving route instructions to each other, humans tend to base those instructions predominantly on *landmarks*, by which we understand distinctive objects in the city environment (Lynch, 1960; Denis et al., 1999). While it is appropriate to give relative directions in certain situations, where such an instruction is unambiguous (Götze and Boye, 2015b), the inclusion of landmarks of vital in more complex navigation situations. It would therefore be desirable if route-giving systems could do the same. In fact, it has been shown that the inclusion of landmarks into system-generated pedestrian routing in-

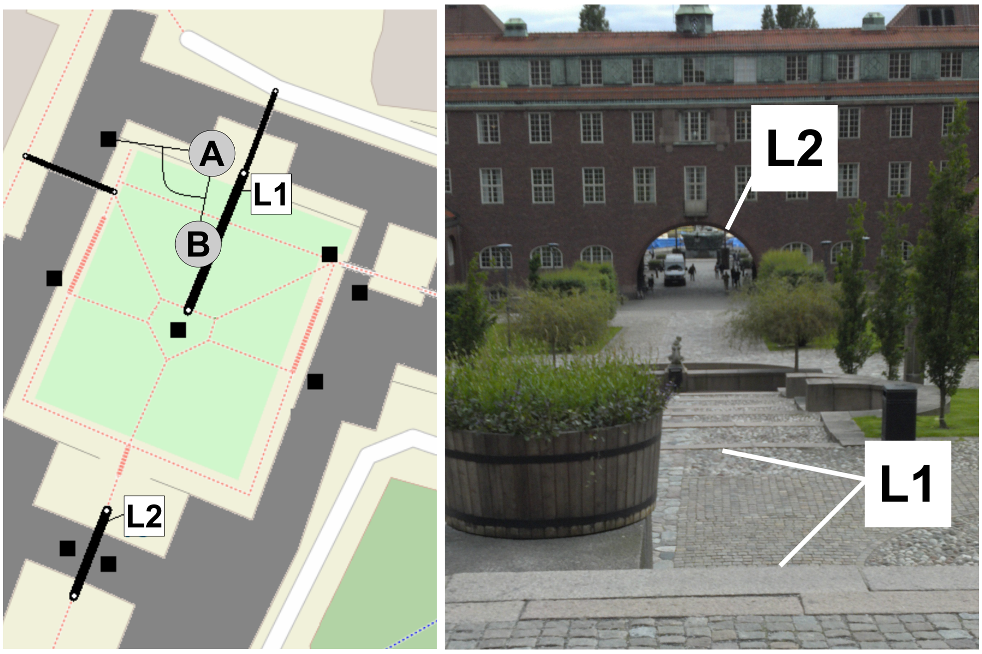

Figure 1: An example segment for the utterance: "I continue in this direction down *the steps* [L1] towards *the arch* [L2]" A and B indicate the start and the goal position respectively. The photo on the right shows the view from the pedestrian's perspective.

structions raises the user's confidence in the system, compared to a system that only gives relative direction instructions (Ross et al., 2004).

However, in each situation there will be a variety of landmarks to choose from, and it is not obvious *which* landmark(s) to include in a particular route instruction. Humans choose objects as landmarks that are *salient* in a particular situation, i.e. that are prominent in a way that makes them easily recognizable. Several researchers have proposed schemes for automatically computing salience values for landmarks (Raubal and Winter, 2002; Duckham et al., 2010; Nothegger et al., 2004). These schemes are typically based on different features that are known to influence salience, like size, visibility and shape, and are intended to be valid for all users. The extent to which each of these features impacts the final salience score is determined by manually setting weights for them, based on different heuristics.

In this article, we take a different approach. Our assumption is that salience is *user-dependent*: different users would find different landmarks to be the most salient in a given situation. Furthermore, our approach is data-driven: Our aim is to (semi-)automatically derive salience measures from examples of users describing the way themselves. We assume that when describing the way, people intuitively select the landmarks they find the most salient in that particular situation. By analyzing and generalizing from such human route descriptions, we aim to construct a mathematical model that can predict salience in new, unseen situations. Note that at this point, we are only interested in the landmarks themselves, not in how to verbalize a reference to them.

As an example, Figure 1 shows a situation with some of the landmarks that could be referred to. Black squares indicate single entities such as shops or entrances to a building, black lines indicate paths, such as streets or stairs, and dark grey shading indicates buildings. In this particular example, the person walking from A to B referred to the landmarks labelled as $L1$ and $L2$: "I continue in this direction down the steps towards the arch". Assuming that these landmarks are the most salient for the user, the system should preferably choose the same landmarks when encountering this situation and thus reduce the cognitive load that is needed to identify a landmark as far as possible.

Note that whenever a person uses a landmark L in a description, he is preferring L over a number of other candidates that *could have been* used in the description but were not. That is to say that L has a higher score according to the person's personal salience model than any other candidate M does. This observation will form the basis of our method, which we will elaborate further in Section 5.

To obtain landmark references that we can learn from, we have performed a study in which users have walked a route in the city of Stockholm, describing the way as they are walking along it. From these descriptions we can obtain information about which landmarks they refer to. An open geographic database (OpenStreetMap, Haklay and Weber, 2008) serves as the basis for computing relevant features. This article extends previous work of ours (Götze and Boye, 2013), where we computed salience models from "armchair" data where users described a route posterior to having walked it. By letting our subjects describe routes as they walk them, as in the study described here, we are aiming to obtain more realistic references and, eventually, better salience models. Our ultimate goal is then to enrich our present system for city navigation (Boye et al., 2014) with personalized salience models.

2. Related Work

Various research has investigated the way in which navigational knowledge is communicated by and to pedestrians by means of natural language (Couclelis, 1996; Denis, 1997; Allen, 1997, 2000; Daniel and Denis, 1998; Denis et al., 1999; Rehrl et al., 2009; Mast et al., 2010). The majority of this research is done on instructions that are given prior to walking. The instruction receiver needs to memorize the turning points and associated actions. This implies a strong need on the instructions to be correct, as well as the turning points to be easily memorizable and recognizable. Landmarks are extensively used to achieve both these needs (Lovelace et al., 1999; Denis, 1997; Denis et al., 1999).

Some research also focusses on guiding visually impaired or disabled pedestrians (Dodson et al., 1999; Helal et al., 2001), whose information needs and ways of communicating the information differ from the results found in other studies.

The focus here is on spoken intructions that are given step by step, while the pedestrian is walking. This allows for possible misunderstandings to be resolved on the spot in an interactive way, as is the long-term goal for our navigation system.

2.1. Landmarks in Pedestrian Navigation

Landmarks are found to play a vital role in both giving and understanding route instructions. They are used to identify points at which actions are to take place, at points where actions could take place, for confirmation along the route, or as general orientation points when they are farther away (Lovelace et al., 1999; Michon and Denis, 2001). Ross et al. (2004) found that they increase the pedestrian's confidence in an automatic system, compared to a system that only gives relative direction instructions. Street names and distance information ("In 200 meters turn into High Street") are dispreferred kinds of information (May et al., 2003; Schroder et al., 2011; Tom and Denis, 2004), they result in more turning errors and lower confidence.

There are several definitions for the term 'landmark', all of which acknowledge an element's prominence in a particular situation and its potential to serve in a cognitive representation of a route (Lynch, 1960; Presson and Montello, 1988; Sorrows and Hirtle, 1999). We are using the term landmark to denote any structure (or set of structures) in the environment of the speaker,[1] such as buildings, areas like parks, shops, paths of any kind, intersections, etc. We are explicitly not excluding streets as landmarks, because Tom and Tversky (2012) have shown that it is not streets per se that are dispreferred as landmarks, but the usage of street names because they can be hard to recognize. This is reflected in our data, in which subjects frequently refer to streets.

2.2. Landmark Salience

When choosing a landmark for use in a route instruction, people do not choose randomly, but try to pick a salient landmark, i.e. a landmark that will be eas-

[1] We leave the incorporation of global landmarks for future research.

ily recognizable (and memorizable in the case of giving instructions prior to walking) for the instruction receiver.

Several kinds of features are found to play a role in determining a landmark's salience, most of them contrast a landmark to its surroundings. The three types of salience features that Sorrows and Hirtle (1999) identify are visual (the landmark stands in visual contrast to its surroundings), structural (the landmark's location is prominent), and cognitive (the landmark's function makes it salient). More recently, efforts have been undertaken to automatically compute the salience of landmarks for given navigation situations.

Raubal and Winter (2002) propose a formal model of landmark salience based on the three types of salience identified by Sorrows and Hirtle (1999). For each type of salience, visual, cognitive, and structural, they propose measures that contribute to it, and properties that describe them. For instance, one measure of visual salience is the façade area of a building, that can be described by its height and width. All measures are weighted and combined into a final salience score by summing them. Except for visibility, which depends on the pedestrian's position, the properties are properties of the landmark itself. A statistical test is proposed to find significant differences between the target landmark and surrounding landmarks, for which they primarily consider buildings. Nothegger et al. (2004) extend this work with an evaluation study in which human subjects are shown panoramic views of intersections and they are asked to choose the most prominent façade. The automatically computed salience measures reflect the human choices, thus proving the suitability of their model.

Duckham et al. (2010) move away from computing the salience of individual landmarks, because the necessary data, such as detailed information about color or shape, is often hard to obtain. They propose to measure salience on the basis of an object's category. They are using a heuristic to determine how suitable a certain category is as a landmark: experts were asked to rate landmark categories according to a set of nine factors that are proposed to describe the salience types of Sorrows and Hirtle (1999). Ratings were given on a five-point scale according to how suitable a specific instance of a category would be as a landmark, and how frequently such an instance occurs. The final score of a category is computed as the weighted sum of these rankings. The landmark categories are manually defined and assumed to be different for different countries. As candidate landmarks a wide range of objects is considered, such as buildings of many kinds, parks, or smaller structures such as mailboxes.

Elias (2003) approaches the task of determining the most salient building of a given set in a different way. She uses semantic features about the buildings' usage and function as well as geometric features reflecting the positioning of the buildings. She applies a clustering algorithm to find a landmark candidate. The approach is based on the idea that a suitable landmark will be an outlier in terms of the used features and not fit into the found clusters. This approach works well for an artificial test dataset.

3. Data Collection

For this study, we asked 6 subjects (4 male, 2 female, average age 28.8) to walk a specific route and describe their path in a way that would make it possible for someone to follow them. Thereby, instead of reading information from a 2-dimensional map, we put the subjects into the environment in which we would later like to guide them, i.e. they can now see the environment in the same way as users of our route-giving system experience it later.

The study was set up as a Wizard-of-Oz study (Dahlbäck and Jönsson, 1989) in which the subjects were asked to describe the way to a spoken dialog system. They were told that the system, like them, had a 3-dimensional and 1st-person view of the environment. The subjects did not receive any particular instructions on how to interact with the system, but were advised to talk in a way they thought was suitable. In this way, all subjects were explaining to the same listener about whom they had no more knowledge than that it was a machine, and we could restrict them somewhat in the way they would formulate their instructions (cf. Kennedy et al., 1988). The role of the experimenter (the "wizard" acting as the machine) was to acknowledge the subjects' descriptions by saying "okay", or asking for a repetition or clarification in the case that there was an interruption in the speech channel, such as too much background noise from the traffic.

The descriptions were collected in English. All subjects reported to be fluent in English. Two of them reported to be only slightly familiar with the area, four reported to be familiar or very familiar. All were able to complete the task.

3.1. Task and Apparatus

The subjects were equipped with an Android mobile phone (Motorola Razr) that ran an application which allowed us to record their GPS coordinates and speech signal (cf. Hill et al., 2012; Boye et al., 2014). It also allowed to send messages from the experimenter to the subject via text-to-speech (TTS). The experimenter sat in a laboratory and used an interface which allowed him to see the subject's position on a map and type messages.

Speech signal and GPS coordinates were automatically logged and time-stamped, thereby allowing to align speech transcriptions with a subject's GPS coordinates. The route that the subjects were asked to walk was a round tour that started and ended outside the doors of our laboratory. The route was

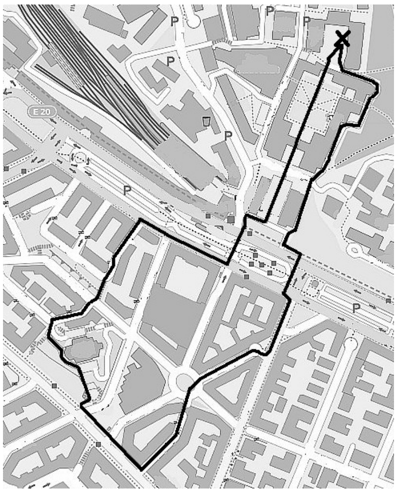

Figure 2: The map of the route that the subjects were asked to follow.

approximately two kilometers long and was given to the subjects on an unlabelled map which is shown in Figure 2, where start and end point are indicated by “X”. The map had street and other names removed, as well as common symbols, e.g. for churches or bus stops.

3.2. Analysis

Two of the subjects deviated slightly from this given route, all others followed the path shown on the map in Figure 2. Subjects could choose in which direction to start the tour, three chose one direction and three the other. The subjects took on average 28 minutes and 25 seconds to complete the tour.

The recorded speech was transcribed and annotated using the Higgins Annotation Tool.[2] Speech segments were annotated as either descriptive (“and then you can see a church to your left”) or instructive (“walk towards the church”). The distinction between an instruction and a description was made based on lexical cues, i.e. the choice of verb, alone. We restrict ourselves here to examining those segments that are instructive, i.e. that specify a movement between a starting point A and a goal point B. The points A and B are GPS coordinates that are derived from the recordings of each subject.

Each of these segments is then annotated with all landmarks from our geographic database that the subject referred to. In the example in Figure 1, the GPS coordinates indicate where the instruction was given and where the next instruction followed. In this example, the subject referred to two objects, “the steps” and “the arch”. Both these objects are entities in the geographic

[2] `http://www.speech.kth.se/hat/`

database, indicated by lines in the figure (cf. Götze and Boye, 2015a, for an overview of the kinds of references that the subjects gave).

4. Problem Encoding

4.1. Learning from Route Segments

For each of our six subjects, we thus have a number of annotated route segments, each describing the path from A (the *starting position* of the segment) to B (the *goal position* of the segment), and at least one landmark that the subject referred to (his *preferred* landmark(s) in this segment). Segments where the subject did not refer to anything at all were excluded from this experiment.

The positions A and B are mapped to their closest nodes in the geographic database. All landmarks in the contexts of these two nodes are considered as possible candidates. A landmark belongs to the *context* of its closest road node. We will refer to this set of landmarks as the *candidate set* for A and B. In Figure 1, a part of this set is visualized as square-shaped icons (for nodes), wide lines (for roads, paths, etc.), or dark grey shading (for buildings). The candidate set for the segment (i.e. all landmarks the user *could have* referred to in a given situation) was automatically computed from the database and contains on average 30 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, or he referred to something that is farther away, and does not belong to the context of neither A nor B. If none of the user-preferred landmarks is part of the candidate set, the route segment was removed from the learning problem.

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.

4.2. OpenStreetMap

For geographic data, we are relying on the OpenStreetMap (OSM) geographic database (Haklay and Weber, 2008). OSM is a freely available crowd-sourced database used in different areas of research, e.g. in robot navigation (Hentschel and Wagner, 2010), in indoor navigation (Goetz, 2012), and in pedestrian navigation (Rehrl et al., 2010). It has two basic data structures:[3] *nodes* and *ways*. Nodes can represent entities in their own right, e.g. intersections, bus stops, or house entrances, but they can also act as the building

[3] We are disregarding OSM *relations* for the time being.

blocks of *ways* (sequences of nodes). Ways are used to represent street segments, buildings, or areas. In what follows, we will avoid the ambiguous term "way", and rather talk about buildings, streets, etc.

OpenStreetMap data is categorized according to an extensive scheme of tags[4] that specifies, for example, how an entity can be represented as a shop, how names are added, or how to indicate speed limits on different parts of a road. Since the data is crowd-sourced on a voluntary basis, it tends to contain inconsistencies in the way tags are applied. Furthermore, the large number of tags results in a different level of detail in different areas and a separation of entities that cognitively belong together, e.g. different segments of the same street are separate entities in OSM (each with its own identifier), because they have different speed limits, or because a bus line is using part of the street. When selecting a landmark from a set of available objects we want to treat such objects as one. In a candidate set, their vectors are therefore combined into one.

4.3. Features

The method described in Section 5 requires every landmark L to which the user can refer to be modelled as a vector of features. In this study, we use a vector of 18 features that are automatically computable, most of them on the basis of the geographic database. Note that we are not making any explicit assumptions about what feature values will positively (or negatively) influence salience. This will instead be reflected in the learned weights.

The following features are used:

Distance

The distances to a landmark are capturing both structural and visual aspects of the scene. Landmarks that are closer to the speaker are more likely to take up a larger field of view. In the case where the landmark is a road or building, distances are computed as the minimum of the distances to each of the nodes that make up the road or building. We are computing the distances:

- between the user's position A and the landmark L (distAL)
- between the landmark L and the goal node B (distLB)

Angle

The angle in which the landmark is located with respect to the pedestrian's walking path gives us structural information. In the case

[4] http://wiki.openstreetmap.org/wiki/Map_Features

where the landmark is a building, the angle is computed as the average of the angles when using each of the nodes in the building to compute the line AL. We are computing the angle:

- between the lines AL and AB (angle)

Name

Whether a landmark has a name or not can be useful information for visual (there is a sign) or cognitive salience (the name is widely known because of the landmark's function). Note that this feature does not reveal whether the landmark was referred to with its name. As mentioned in Section 3, the subjects believed that they were talking to an automatic system and they may have chosen to use the landmark's category instead. The value is assigned as follows:

- The categorial attribute name has 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 a street, and the value 0 otherwise.

Type

Landmarks in our data are often referred to by their type, around 97% of all referring expressions contain a type specification (cf. Götze and Boye, 2015a). Following the approach of Duckham et al. (2010), a landmark's type (or category) contributes to its salience by the type's frequency and its general suitability as a landmark. Type values are assigned as follows:
Each landmark is of at least one type, which is indicated by the value 1 in the corresponding slot. In OpenStreetMap, entries are annotated with types in the form of tags, which we summarize into the following type features:

road if the landmark is a road or part of a road. In OSM, roads are tagged as highway and have different values depending on their function. This feature includes all different kinds of streets, or paths of any kind. In OSM, these entities can have any of the following values for the tag highway: *primary*, *secondary*, *tertiary*, *motorway*, *residential*, *footway*, *cycleway*, *path*, *pedestrian*, or *steps*.

building if the landmark has the OSM tag building, indicating a building of any sort. A building can have a special type (buildingSpec), such as a *hospital*, *school*, or *church*, or it has no special function (buildingGen), such as a residential building.

eating for a *restaurant* or *cafe*, etc.

leisure summarizes entities tagged as e.g. *theater*, *library*, or *museum*.
shop a *supermarket*, a *pharmacy*, or other entities tagged as *shop*.
entrance for a specific street address.
area for a *park* or a *construction site*, etc.
structure for a *statue* or a *fountain*, etc.
crossing for a pedestrian crossing with or without traffic lights
pubTrans for an entity belonging to the public transportation system, e.g. a *bus stop*.
other for other identifiable entities, e.g. *bench* or *bicycle parking*.

Extending the OSM Features

We are extending this feature set with the following features:

- The feature duplicates counts how many other objects there are in the candidate set that have the same values for all type features. The intuition is that if there are several objects of the same type, more effort is needed to distinguish one from the other because none of them can be described unambiguously in a simple way. This may play a role in deciding whether to refer to the landmark in question.
- The cognitive feature onRoute applies to objects that belong to the route that the pedestrian is following. It has the value 1 if the object belongs to the route, and 0 if it does not, e.g. a street that the pedestrian crosses but does not walk along. In an automatic navigation system, this feature is available as part of the planned route.

In the example in Figure 1, the stairs that the user refers to (the wider line close to the goal point B), is represented by the vector $(3, 2, 27, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 2, 1)$. 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 succeeding slot indicates that the landmark does not have a name. The values in the next 12 slots indicate that the landmark is a kind of road, but no other type. The two final slots indicate that there are two other entities in the candidate set that are of the same kind (having a value of 1 only for the road feature), and that the entity is part of the route that the subject walks along.

All non-categorical features are normalized within the segment in which they appear. Distances and angles, as well as the duplicates feature that counts types, are then relative between the landmarks that appear in the same segment. The landmark that is farthest away from the speaker will have the value 1 for the feature distAL.

5. Salience Models

Previously we noted that whenever a person uses a landmark L in a description, he is preferring L over a number of other candidates that *could have been* used in the description but were not. That is to say that the person (probably unconsciously) finds L more salient than any other available candidate M. Our goal is now to create a mathematical model of salience that generalizes from these observations. This model can then be used to select a suitable landmark to use in routing instructions in new, hitherto unseen situations.

First, note that the available data can *not* be interpreted as a measure of *absolute* salience. The preferred landmark L might be perceived as very salient or perhaps not very salient at all; all we know is that it is *more* salient than the other available candidates. Therefore it would be inappropriate to, say, use a binary classification method where L is tagged as 'salient' and the other candidates as 'not salient'. Rather, we want to create a model that *ranks* the landmarks from 'best' to 'worst'. Such a model will attach a numerical score to each available landmark indicating its salience, and the landmark with the highest score is considered to be the most salient one. However, it should be emphasized that the numbers themselves are unimportant; they are just a means to get to the ranking, and the numbers do not represent salience in any absolute way. In particular, we cannot compare salience scores between different situations.

For learning such ranked salience models, we use the Ranking SVM Algorithm described by Joachims (2002). This algorithm has been used for various non-linear ranking tasks, e.g. in Named Entity Recognition (Bunescu and Paşca, 2006) and Sentiment Classification (Kennedy and Inkpen, 2006).

As described in the previous section, each landmark can be represented as a vector of numerical features, $\mathbf{x} = (x_1, \ldots, x_n)$ specifying scores 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 $s(\mathbf{x})$ of a landmark 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 knows that such a model exists. Instead we automatically infer the model as follows:

When a person uses a landmark L in a description rather than landmark M, we can represent this as the inequality $\mathbf{w} \cdot (\mathbf{x_L} - \mathbf{x_M}) > 0$, where $\mathbf{x_L}$ and $\mathbf{x_M}$ are the vectors representing L, and M, respectively. This inequality expresses the fact that L is more salient than M according to the model represented by $\mathbf{w}$. Each route description from the user involving a landmark thus generates a number of inequalities. Let m be the total number of inequalities for all route segment descriptions. Then we want to find a weight vector **w** such that $\mathbf{w} \cdot (\mathbf{x}_{\mathbf{L_i}} - \mathbf{x}_{\mathbf{M_i}}) > 0$, for $1 \leq i \leq m$. (For brevity, we will use the notation $\mathbf{d_i}$ for the difference $\mathbf{x}_{\mathbf{L_i}} - \mathbf{x}_{\mathbf{M_i}}$). Our goal is to find appropriate values for the weights in **w** that satisfy as many of the inequalities $\mathbf{w} \cdot \mathbf{d_i} > 0$ as possible.

This can be done by solving the following optimization problem:

$$\begin{array}{lll}
\text{minimize} & \frac{1}{2}\mathbf{w} \cdot \mathbf{w} + c \sum_{i=1}^{m} \xi_i & \\
\text{where} & \mathbf{w} \cdot \mathbf{d_i} + \xi_i \geq 1, & i = 1 \ldots m \\
 & \xi_i \geq 0, & i = 1 \ldots m
\end{array}$$

Assuming that a person is not always consistent in his preferences, this formulation of the problem introduces slack variables ξ_i and adds a penalty c on those variables (see Joachims, 2002, 2006, for details).

6. Results

Recall that an *instance* for our ranking problem is a candidate set together with one or several preferred landmarks (see Section 4.1), that give rise to a number of inequalities as explained above. For evaluation, the set of all instances for a particular subject was split into a training set and a test set. The training set, two thirds of the segments, was used to derive a salience model **w** according to the method presented in Section 5. This was repeated several times for different permutations of route segments and averages were computed. To evaluate **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 user-preferred landmarks had the best salience according to the model **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.

The results are presented in Table 1 and show the following evaluation measures:

FHS First Hit Success is the proportion of route segments in which a user-preferred landmark was ranked highest by the inferred model, i.e. the proportion of successful instances.

MRR Mean Reciprocal Rank (cf. Radev et al., 2002): If a user-preferred landmark is ranked as the nth landmark by the inferred model, its reciprocal rank is 1/n. The total reciprocal rank is the sum of the reciprocal ranks of all user-preferred landmarks in the segment. For the mean, this number is divided by the number of user-preferred landmarks.

Ranking The above measures do not account for the total number of landmarks that were available in a particular situation. We therefore intro-

Table 1: Evaluation measures for the derived salience models: First Hit Success (FHS), Mean Reciprocal Rank (MRR), Ranking, and subjects' self-rated familiarity with the area. The numbers represent averages obtained using three-fold cross-validation.

Subject ID	# Segments (total)	FHS	MRR	Ranking	Familiarity (1-6, 6=max)
A	32	**0.55**	0.60	0.13	5
B	34	0.50	**0.66**	**0.05**	6
C	37	0.40	0.59	0.11	3
D	36	0.35	0.51	0.14	6
E	35	0.30	0.53	0.22	2
F	27	0.47	0.52	0.19	6
average		0.42	0.57	0.14	
A+B+C+D+E+F	201	0.35	0.52	0.09	
A+B+C+D+E+F (training on 25 instances)		0.38	0.52	0.14	

duce this measure that reflects the proportion of landmarks that the inferred model ranked higher than the best-ranked user-preferred landmark. Recall that the mean number of available landmarks in a route segment is 30.

In between 30% and 55% of the instances, the inferred salience models rank a user-preferred landmark highest. On average, the user-preferred landmark is among the the 14% highest ranked landmarks. The mean reciprocal rank is on average 0.57.

Table 2 shows two example feature vectors, i.e. two salience models, sorted by the values of the weights. These weights were obtained when training on all instances of subjects A and B, respectively. The different orderings of the features reflects different preferences of these two subjects when choosing landmarks to refer to. For example, subject A seems to preferably choose landmarks that are close to the goal position of the route segment (feature distLB), while the same feature has a value close to zero for subject B, meaning that it does not play a significant role for ranking landmarks.

7. Discussion

The SVM Ranking method manages to mimic the user's salience preferences in 42% of the tested instances. Recall that an instance contained on average 30 landmarks.

How good is this result? Recall that we are aiming for an interactive guid-

Table 2: Comparing the feature weights for two subjects' models (Subjects A and B)

Subject A		Subject B	
Feature	Weight	Feature	Weight
distLB	-1.927	entrance	-0.835
distAL	-1.269	eat	-0.712
angle	-1.039	distAL	-0.651
structure	-0.477	area	-0.575
other	-0.464	buildingSpec	-0.542
pubTrans	-0.456	road	-0.452
shop	-0.389	structure	-0.439
eat	-0.376	angle	-0.363
duplicates	-0.180	crossing	-0.353
buildingSpec	0.000	shop	-0.265
leisure	0.000	pubTrans	-0.088
buildingGen	0.193	distLB	-0.011
road	0.193	other	0.000
entrance	0.430	leisure	0.000
area	0.550	buildingGen	0.090
name	0.810	duplicates	0.437
crossing	0.810	onRoute	1.110
onRoute	1.288	name	1.388

ing scenario, where the system has the option of first confirming with the user that he can identify the landmark, before using it in an instruction. Moreover, since all available landmarks are ranked, the system can use the next-best ranked landmark if the user is unable to recognize the top-ranked one. Another possibility would be for the system to change to a different navigation strategy, such as asking the user to identify what he can see. Such information could be used to further tune the "personal" weights of this user.

We can see that for some users, the ranking produces better results than for others and this seems to be unrelated to the amount of available training data (which was two thirds of the total number of segments). For example, subject E's models were successful in 30% of the test instances. On average, the learned ranking function ranks the landmark that was preferred by E among the 22% highest ranked landmarks in a situation. For subject D, where a similar number of training instances was available, the method achieved a FHS rate of 35% and the preferred landmark was on average among the 14% highest ranked landmarks. A possible explanation for this is that a subject might have changed his strategy for choosing landmarks along the route, thus introducing more inconsistencies when evaluating the set of references as a whole.

Such a change could depend on a (perceived) change of environment, e.g. by entering an unknown area where the pedestrian has to rely more on visual features while in familiar situations he can refer to familiar places by their name. As a reference, we are reporting the subjects' scores of overall familiarity with the area in Table 1.

Table 2 shows salience models of two subjects that differ in which of the features contribute most to a ranking, suggesting that the models should indeed be computed per person rather than having only one model for all. In order to further assess whether a combined model, containing landmark preferences from several subjects, can be useful instead of personal models, we also built such a model. The lower part of Table 1 shows the evaluation measures for both training a model on two thirds of all available data, as well as training on a set that is comparable in size to the personal models (25 training instances). When training on a comparable size of instances as for the personal models, we can see that the combined model does not perform better than the individual models. When increasing the training set size to ca. 134 instances (two thirds), we can see no improvement, which strengthens the plausibility of a personal salience model for each user.

8. Conclusion

We have presented an approach to learn individual salience models for landmarks that are used in navigation instructions, using landmark features that are computable in real-time from crowd-sourced, readily available data. Instead of hand-tuning the weights in a salience function, we are learning a weight model that is individual to each of our subjects and reflects the contribution of different features in selecting a landmark in a given situation.

The evaluation of these models shows promising results. When ranking the available landmarks in a navigation situation, they can often predict the landmark that was chosen by the user and generally ranks the user-preferred landmark high. While the overall results still leave room for improvement, we believe that the described ranking method will be a useful addition to existing methods that compute salience on a variety of features. As discussed in Section 2.2, several methods use weights to account for the impact of different salience features. These weights are hand-tuned on the basis of theoretical research about salience (e.g. Raubal and Winter, 2002). The ranking method we propose allows to learn these weights from data, e.g. from landmark information as collected in a recently developed application by Wolfensberger and Richter (2015). Note that instead of user preference ratings it would also be possible to learn from data that a deployed system collects: the system can collect information about which landmarks worked well in a situation (and should be ranked higher), and which ones did not (and should be ranked lower than all others).

The features we used in this work are simple features that can be easily computed from OpenStreetMap. However, the features are independent of how a landmark was referred to. Only the geographic representation is taken into account, regardless of whether the corresponding feature was also mentioned in the description. For example, an object can be a building or have a name without the reference containing the word "building" or the name of the object. Likewise, the subjects mention features that we currently cannot compute from the OSM database, such as size ("the smaller fountain"), color ("a yellow building"), material ("a brick building"), or slope ("a slight incline"). We plan to further investigate how the features mentioned by the describer can be used in computing salience.

The next step in this work will be to incorporate the learned models in our pedestrian navigation system, and try them out on new situations.

References

Allen, G. L. (1997). From knowledge to words to wayfinding: Issues in the production and comprehension of route directions. In *COSIT*, volume 1329 of *LNCS*, pages 363–372. Springer.

Allen, G. L. (2000). Principles and practices for communicating route knowledge. *Applied Cognitive Psychology*, 14(4):333–359.

Boye, J., Fredriksson, M., Götze, J., Gustafson, J., and Königsmann, J. (2014). Walk this way: Spatial grounding for city exploration. In *Natural Interaction with Robots, Knowbots and Smartphones*, pages 59–67. Springer.

Bunescu, R. and Paşca, M. (2006). Using encyclopedic knowledge for named entity disambiguation. *Proc. of the 11th Conference of the European Chapter of the Association for Computational Linguistics (EACL-06)*, pages 9–16.

Couclelis, H. (1996). Verbal directions for way-finding: Space, cognition, and language. In *The Construction of Cognitive Maps*, volume 32 of *GeoJournal Library*, pages 133–153. Springer.

Dahlbäck, N. and Jönsson, A. (1989). Empirical studies of discourse representations for natural language interfaces. In *Proc. of the 4th Conference of the European Chapter of the Association for Computational Linguistics (EACL-89)*, pages 291–298. ACL.

Daniel, M.-P. and Denis, M. (1998). Spatial descriptions as navigational aids: a cognitive analysis of route directions. *Kognitionswissenschaft*, 7:45–52.

Denis, M. (1997). The description of routes: A cognitive approach to the production of spatial discourse. *Current Psychology of Cognition*, 16(4):409–458.

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*, 13(2):145–174.

Dodson, A. H., Moon, G. V., Moore, T., and Jones, D. (1999). Guiding blind pedestrians with a personal navigation system. *Journal of Navigation*, 52:330–341.

Duckham, M., Winter, S., and Robinson, M. (2010). Including landmarks in routing instructions. *J. Location Based Services*, 4(1):28–52.

Elias, B. (2003). Extracting landmarks with data mining methods. In *COSIT*, volume 2825 of *LNCS*, pages 375–389. Springer.

Goetz, M. (2012). Using crowdsourced indoor geodata for the creation of a three-dimensional indoor routing web application. *Future Internet*, 4(2):575–591.

Google (2013). Google maps navigation.

Götze, J. and Boye, J. (2013). Deriving salience models from human route directions. In *Workshop on Computational Models of Spatial Language Interpretation and Generation 2013 (CoSLI-3)*, pages 36–41.

Götze, J. and Boye, J. (2015a). Resolving spatial references using crowdsourced geographical data. In *Proc of the 20th Nordic Conference of Computational Linguistics, NODALIDA*, pages 61–68. Linköping University Electronic Press.

Götze, J. and Boye, J. (2015b). "Turn Left" Versus "Walk Towards the Café": When Relative Directions Work Better Than Landmarks. In *AGILE 2015*, Lecture Notes in Geoinformation and Cartography, pages 253–267. Springer.

Haklay, M. and Weber, P. (2008). Openstreetmap: User-generated street maps. *Pervasive Computing, IEEE*, 7(4):12–18.

Helal, A., Moore, S., and Ramachandran, B. (2001). Drishti: an integrated navigation system for visually impaired and disabled. In *Proc. Fifth International Symposium on Wearable Computers*, pages 149–156.

Hentschel, M. and Wagner, B. (2010). Autonomous robot navigation based on openstreetmap geodata. In *Intelligent Transportation Systems (ITSC), 2010 13th International IEEE Conference on*, pages 1645–1650.

Hill, R., Götze, J., and Webber, B. (2012). Final data release, wizard-of-oz (woz) experiments.

Janarthanam, S., Lemon, O., Liu, X., Bartie, P., Mackaness, W., Dalmas, T., and Goetze, J. (2012). Integrating location, visibility, and question-answering in a spoken dialogue system for pedestrian city exploration. In *Proc. of the 16th Workshop on the Semantics and Pragmatics of Dialogue (Semdial)*.

Joachims, T. (2002). Optimizing search engines using clickthrough data. In *Proc. of the ACM Conference on Knowledge Discovery and Data Mining (KDD)*, pages 133–142.

Joachims, T. (2006). Training linear svms in linear time. In *Proc. of the ACM Conference on Knowledge Discovery and Data Mining (KDD)*, pages 217–226.

Kennedy, A. and Inkpen, D. (2006). Sentiment classification of movie reviews using contextual valence shifters. *Computational Intelligence*, 22(2):110–125.

Kennedy, A., Wilkes, A., Elder, L., and Murray, W. S. (1988). Dialogue with machines. *Cognition*, 30(1):37–72.

Lovelace, K. L., Hegarty, M., and Montello, D. R. (1999). Elements of good route directions in familiar and unfamiliar environments. In *Spatial Information Theory. Cognitive and Computational Foundations of Geographic Information Science*, volume 1661 of *LNCS*, pages 65–82. Springer.

Lynch, K. (1960). *The image of the city*. The M.I.T. Press, Cambridge, MA.

Mast, V., Smeddinck, J., Strotseva, A., and Tenbrink, T. (2010). The impact of dimensionality on natural language route directions in unconstrained dialogue. In *Proc of the 11th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, SIGDIAL '10, pages 99–102. ACL.

May, A. J., Ross, T., Bayer, S. H., and Tarkiainen, M. (2003). Pedestrian navigation aids: information requirements and design implications. *Personal and Ubiquitous Computing*, 7(6):331–338.

Michon, P.-E. and Denis, M. (2001). When and why are visual landmarks used in giving directions? In *Spatial Information Theory*, volume 2205 of *LNCS*, pages 292–305. Springer.

Nothegger, C., Winter, S., and Raubal, M. (2004). Selection of salient features for route directions. *Spatial Cognition & Computation*, 4(2):113–136.

Presson, C. C. and Montello, D. R. (1988). Points of reference in spatial cognition: Stalking the elusive landmark. *British Journal of Developmental Psychology*, 6:378–381.

Radev, D. R., Qi, H., Wu, H., and Fan, W. (2002). Evaluating web-based question answering systems. In *LREC*. European Language Resources Association.

Raubal, M. and Winter, S. (2002). Enriching wayfinding instructions with local landmarks. In *Geographic Information Science*, volume 2478 of *LNCS*, pages 243–259. Springer.

Rehrl, K., Häusler, E., and Leitinger, S. (2010). Gps-based voice guidance as navigation support for pedestrians, alpine skiers and alpine tourers. In *Proc. of Workshop on Multimodal Location Based Techniques for Extreme Navigation*, pages 13–16.

Rehrl, K., Leitinger, S., Gartner, G., and Ortag, F. (2009). An analysis of direction and motion concepts in verbal descriptions of route choices. In *Spatial Information Theory*, volume 5756 of *LNCS*, pages 471–488. Springer.

Ross, T., May, A. J., and Thompson, S. (2004). The use of landmarks in pedestrian navigation instructions and the effects of context. In *Mobile HCI*, volume 3160 of *LNCS*, pages 300–304. Springer.

Schroder, C. J., Mackaness, W. A., and Gittings, B. M. (2011). Giving the 'right' route directions: The requirements for pedestrian navigation systems. *Transactions in GIS*, 15(3):419–438.

Sorrows, M. E. and Hirtle, S. C. (1999). The nature of landmarks for real and electronic spaces. In *COSIT*, volume 1661 of *LNCS*, pages 37–50. Springer.

Tom, A. and Denis, M. (2004). Language and spatial cognition: comparing the roles of landmarks and street names in route instructions. *Applied Cognitive Psychology*, 18(9):1213–1230.

Tom, A. C. and Tversky, B. (2012). Remembering routes: Streets and landmarks. *Applied Cognitive Psychology*, 26(2):182–193.

Wolfensberger, M. and Richter, K.-F. (2015). A mobile application for a user-generated collection of landmarks. In *Web and Wireless Geographical Information Systems*, volume 9080 of *LNCS*, pages 3–19. Springer.