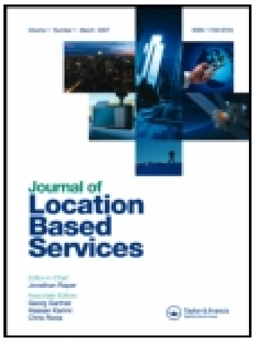

Taylor & Francis
Taylor & Francis Group

Journal of Location Based Services

ISSN: 1748-9725 (Print) 1748-9733 (Online) Journal homepage: http://www.tandfonline.com/loi/tlbs20

Learning landmark salience models from users' route instructions

Jana Götze & Johan Boye

To cite this article: Jana Götze & Johan Boye (2016) Learning landmark salience models from users' route instructions, Journal of Location Based Services, 10:1, 47-63, DOI: 10.1080/17489725.2016.1172739

To link to this article: http://dx.doi.org/10.1080/17489725.2016.1172739

Published online: 22 Apr 2016.

JOURNAL OF LOCATION BASED SERVICES, 2016
VOL. 10, NO. 1, 47–63
http://dx.doi.org/10.1080/17489725.2016.1172739

Taylor & Francis
Taylor & Francis Group

Learning landmark salience models from users' route instructions

Jana Götze and Johan Boye

School of Computer Science and Communication, KTH Royal Institute of Technology, Stockholm, Sweden

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 modelled 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.

ARTICLE HISTORY
Recevied 11 December 2015
Accepted 17 March 2016

KEYWORDS
Pedestrian navigation; landmark salience; ranking algorithm; spoken route descriptions

1. Introduction

Recently there has been an increasing interest in systems capable of providing natural language route instructions to pedestrians in a city environment (Boye et al. 2014; Google Inc. 2013; Janarthanam et al. 2012; Rehrl, Häusler, and Leitinger 2010). 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, minimising 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 (Denis et al. 1999; Lynch 1960). While it is appropriate to give relative directions ('Turn left/right') in certain situations, where such an instruction is unambiguous (Götze and Boye 2015b), the inclusion of landmarks is 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 instructions raises the user's confidence in the system, compared to a system that only gives relative direction instructions (Ross, May, and Thompson 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

CONTACT Jana Götze ✉ jagoetze@kth.se

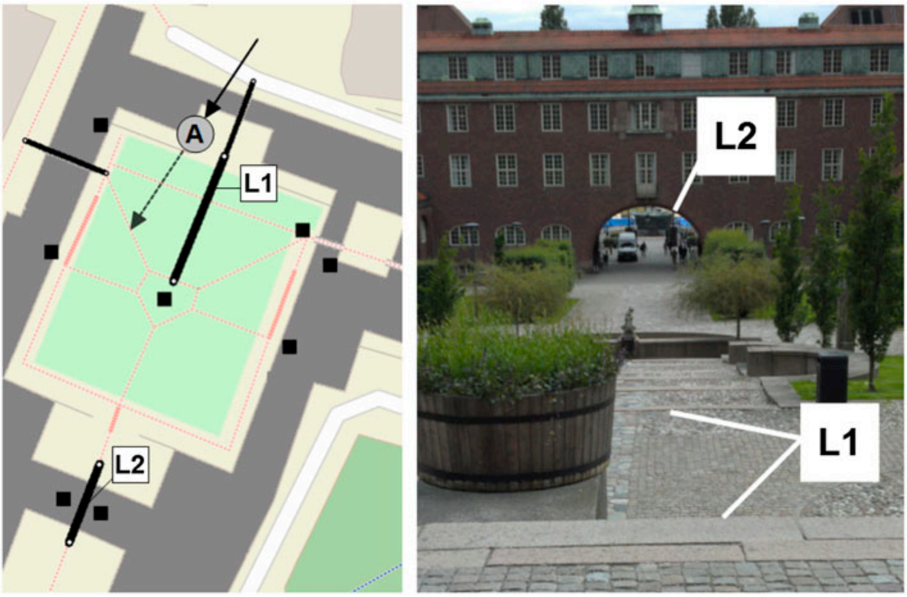

Figure 1. An example segment for the utterance: 'I continue in this direction down *the steps* [L1] towards *the arch* [L2]' A indicates the pedestrian's position.
Note: The arrows indicate the pedestrian's walking direction. The photo on the right shows the view from the pedestrian's perspective.

way that makes them easily recognisable. Several researchers have proposed schemes for automatically computing salience values for landmarks (Duckham, Winter, and Robinson 2010; Nothegger, Winter, and Raubal 2004; Raubal and Winter 2002). 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, pedestrians intuitively select the landmarks they find the most salient in that particular situation. By analysing and generalising 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 verbalise 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 *L*1 and *L*2: '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*. This observation will form the basis of our method, which we will explain in Section 5.

To obtain landmark references that we can learn from, we have performed a study in which pedestrians 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 [OSM], 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 'arm-chair' data where pedestrians 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 personalised 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 (Allen 1997, 2000; Couclelis 1996; Daniel and Denis 1998; Denis 1997; Denis et al. 1999; Mast et al. 2010; Rehrl et al. 2009). The majority of this research is done on instructions that are given prior to walking. The instruction receiver needs to memorise 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 memorisable and recognisable. Landmarks are extensively used to achieve both these needs (Denis 1997; Denis et al. 1999; Lovelace, Hegarty, and Montello 1999). Some research also focuses on guiding visually impaired or disabled pedestrians (Dodson et al. 1999; Helal, Moore, and Ramachandran 2001), whose information needs and ways of communicating the information differ from the results found in other studies.

The focus here is on spoken instructions 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, Hegarty, and Montello 1999; Michon and Denis 2001). Ross, May, and Thompson (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, Mackaness, and Gittings 2011; Tom and Denis 2004), as they result in more turning errors and lower confidence.

There are several definitions of 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 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 recognise. 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 easily recognisable (and memorisable 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 in 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. They propose to use a statistical test to find significant differences between the target landmark and surrounding landmarks, for which they primarily consider buildings. Nothegger, Winter, and Raubal (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 showing the suitability of their model.

Duckham, Winter, and Robinson (2010) move away from computing the salience of individual landmarks, because the necessary data, such as detailed information about color or shape, are 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. A wide range of objects is considered as candidate landmarks, 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 position 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 will not fit into the found clusters. This approach works well for an artificial test data-set.

3. Data collection

For this study, we asked 10 subjects (9 male, 1 female, average age 27.3) 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 two-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 dialogue system. They were told that the system, like them, had a three-dimensional and first-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 (cf. Figure 3 in Section 6). 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. Boye et al. 2014; Hill, Götze, and Webber 2012). It also allowed to send messages from the experimenter to the subject via text-to-speech. 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 department. The route was approximately two kilometers long and was given to the subjects on an unlabelled map which is shown in Figure 2, where start and end points are indicated by 'X'. The map had street and other names removed, as well as common symbols, e.g. for churches or bus stops, in order to force the subjects to rely on information that they could see in their physical environment rather than on the map.

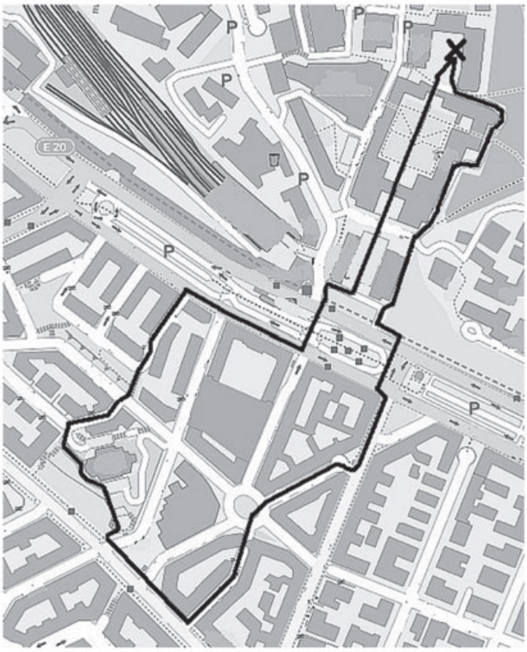

Figure 2. The map of the route that the subjects were asked to follow.

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, six chose one direction and four the other. The subjects took on average 31 min and 34 sec to complete the tour.

The recorded speech was segmented, transcribed and annotated using the Higgins Annotation Tool.[2] Each segment constitutes a new route situation. The pedestrian's position *A* is a GPS coordinate derived from the corresponding recording.

Each of these segments is annotated with all landmarks that the subject referred to. In the example in Figure 1, the GPS coordinate indicates where the utterance was made. In this example, the subject referred to two objects, 'the steps' and 'the arch' (for an overview of the kinds of references that the subjects gave cf. Götze and Boye 2015a).

4. Problem encoding

4.1. Learning from route segments

For each of our 10 subjects, we thus have a number of annotated route situations, each describing the position at which the subject is located, and at least one landmark that the subject referred to (his *preferred* landmark(s) in this situation). Situations where the subject did not refer to anything at all were excluded from this study.

From the pedestrian's position *A*, we compute the *candidate set*, all landmarks in the vicinity that the pedestrian *could have* referred to in a given situation. In Figure 1, a part of this set is visualised as square-shaped icons (for nodes), wide lines (for roads, paths, etc.)

or dark grey shading (for buildings). The candidate set for the segment was automatically computed from the database and contains on average 33 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 the subject's position. 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 a geographic representation of the city, we are relying on the OSM geographic database (Haklay and Weber 2008). OSM is a freely available crowd-sourced database used in many areas of research, e.g. in robot navigation (Hentschel and Wagner 2010), in indoor navigation (Goetz 2012) and in pedestrian navigation (Rehrl, Häusler, and Leitinger 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 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.

OSM data is categorised 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 are crowd-sourced on a voluntary basis, it tends to contain inconsistencies in the way tags are used. 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 representations 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 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:

4.3.1. Positional features

- *Distance* The distance from the pedestrian to a landmark is 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.

- *Angle* The angle in which the landmark is located with respect to the pedestrian's previous walking path gives us structural information. In the case 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. The values for this feature range between 0 ('straight') and 180 ('behind'). 'Left' and 'right' are collapsed, a value of 20 can mean left or right.

4.3.2. Type features

The type features are binary features. An entity either is of a certain type and has value 1 for this feature, or it has value 0 if it is not of that type. As type features, we are using the full OSM tag set from our city model and are referring to wiki.openstreetmap.org/wiki/Map_Features[5] as 'the (wiki) specification'. OSM tags are used in the following way: Each tag is split into its tag key and tag value, each is its own binary feature, with the following exceptions:

(1) For tag keys with values that are specific for each entity, such as *name*, *website*, *wikipedia*, *opening_hours*, only the tag key is added as a feature (It is meaningful to know that an entity has a name, but not its specific value).

(2) For tags that have binary values (*yes*/*no*), tag key and tag value are merged into one feature. For example, `<tag k="steps" v="yes"/>` becomes *steps:yes*.

(3) Some tags are excluded altogether. These tags are specifying information that is not relevant for pedestrians (e.g. *maxspeed*, *oneway*), or not relevant for the task (e.g. *source*). Tags are excluded on the basis of their description in the wiki specification. That means that other, user-defined tags that also carry non-relevant information, are not excluded from the feature set (see the discussion in Section 4.3.5).

The obtained set is then further processed:

(4) If there are two features *f* and *f:yes*, they are merged into one. This situation can arise when parts of a tag are used in different ways, e.g. for the tags `<tag k="highway" v="steps"/>` and `<tag k="steps" v="yes"/>`, which become the features *highway*, *steps*, and *steps:yes*. The latter two express the same concept, so they are merged to avoid duplication.

This procedure amounts to 427 type features.

4.3.3. Context 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.

4.3.4. Example landmark representation

Consider again the route situation shown in Figure 1. Landmark *L*2 ('the arch'), has the following set of tags in the city model:

```
<tag k="highway" v="footway"/>
```

```
<tag k="layer" v="-1"/>
<tag k="source" v="yahoo; survey"/>
<tag k="tunnel" v="yes"/>
```

The *distance* feature will have a value of 6 (the 2-logarithm of the distance to the closest node of the object). The *angle* will be 13.62. The object will have a value of 1 for the type features *highway*, *footway*, *layer* and *tunnel:yes*, and 0 for all other type features. The value for *duplicates* is 0, because there are no other objects available as landmarks in this situation that have the same type feature specification.

4.3.5. Discussion of the type features

OSM tags are crowd-sourced. They follow a certain specification, but rules are sometimes interpreted in different ways, and nothing prevents a contributor from adding their own tags. A way to ensure more consistency and less noise in the data is to create some intermediate categories that function as bins for several OSM tags that express similar concepts, e.g. a *street* feature that collects all entities that have the *highway* key.

This however requires a great amount of pre-processing to ensure that all possible tags are covered and sorted into the correct bin. Bins could be created on the basis of the wiki specification, which would exclude many user-defined tags that contain possibly useful information. On the other hand, creating bins on the basis of the specific city model (including user-defined tags) requires repeating the process for each new data-set. Such bins also do not completely remove noise, as some entities will always be tagged incorrectly.

We have previously created such bins based on our data (cf. Götze and Boye 2013), but are now suggesting to skip this step for the above named reasons. The above processing steps are simpler and can be automatically applied. Comparing the results for a subset of our experiment data, we also get better results for this higher dimensional data (427 instead of 8 type features).

Instead, we use whatever tags there are available in the city model with only few exceptions but including user-defined tags. The steps that do require manual processing are steps 1 and 3 (deciding from which tags only to use the key and which tags to exclude altogether). The definition of which tags fall into these categories can however be re-used for new data-sets. Note also that we are not using the complete set of distinct OSM tag keys, which is currently (Oct 2015) larger than 56, 000. Those tags that do not occur in our study area would have the same value for all objects (namely 0) and thus not influence the result in any way.

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 generalises 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 emphasised 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 $\mathbf{w}$ such that $\mathbf{w} \cdot (\mathbf{x_{L_i}} - \mathbf{x_{M_i}}) > 0$, for $1 \leq i \leq m$. (For brevity, we will use the notation $\mathbf{d_i}$ for the difference $\mathbf{x_{L_i}} - \mathbf{x_{M_i}}$). Our goal is to find appropriate values for the weights in $\mathbf{w}$ that satisfy as many of the inequalities $\mathbf{w} \cdot \mathbf{d_i} > 0$ as possible.

This can be done by solving the following optimisation problem:

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

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, we perform fivefold cross-validation on the set of all instances for a particular subject. Each training set was used to derive a salience model $\mathbf{w}$ according to the method presented in Section 5. To evaluate $\mathbf{w}$, the salience of each

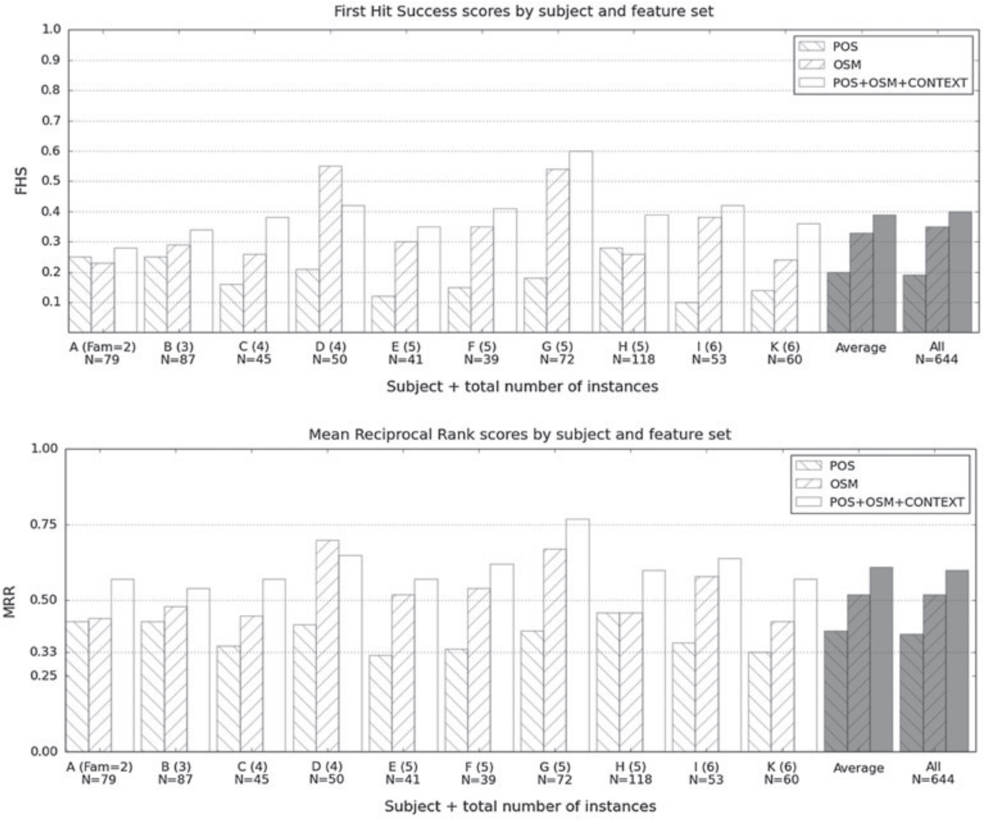

Figure 3. Evaluation measures for the derived salience models: First Hit Success (FHS), Mean Reciprocal Rank (MRR).

Note: The numbers represent averages obtained using fivefold cross-validation. 'Average' are averages over all subjects. 'All' are measures for combined models, i.e. not distinguishing between subjects. Numbers in parentheses are self-rated familiarity scores (max=6). POS uses only positional features, OSM only type features. OSM+POS+CONTEXT uses all features.

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 results are presented in Figure 3 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.

On average, in 39% of the instances, the inferred salience models rank a user-preferred landmark highest. The mean reciprocal rank is on average 0.61, which can be interpreted as 'an average rank better than 2'. As a baseline for comparison, the figure also shows results

Table 1. Comparing the feature order for two subjects' models (Subjects G and H): The 10 highest and lowest weights for the type features and weights for positional and context features.

Subject D		Subject G	
Weight	Feature	Feature	Weight
1.529	lit:yes	steps:yes	1.383
1.499	name	secondary	0.914
0.901	housenumber	fountain	0.768
0.901	street	name	0.738
0.709	service	layer	0.603
0.633	residential	tunnel:yes	0.603
0.624	unclassified	residential	0.596
0.429	secondary	education	0.504
0.418	lane	christian	0.449
0.396	lcn:yes	church	0.449
...			...
−0.303	psv:yes	ramp:no	−0.278
−0.309	operator	motorcar:no	−0.343
−0.341	up	website	−0.349
−0.368	footway	unclassified	−0.377
−0.377	amenity	hospital	−0.390
−0.379	website	junction	−0.456
−0.503	layer	roundabout	−0.456
−0.503	tunnel:yes	track	−0.491
−0.540	junction	psv:yes	−0.612
−0.540	roundabout	secondary_link	−0.624
−0.395	distance	distance	−0.495
0.001	angle	angle	−0.007
0.015	duplicates	duplicates	0.012

for using only the positional features *distance* and *angle* (POS, cf. Section 4.3.1) and for only using the type features (OSM, cf. Section 4.3.2).[6] Recall that choosing at random means choosing from on average 33 landmarks.

We can see that for most subjects, combining the features and including the context feature (POS+OSM+CONTEXT) improve the outcome of the ranking, the only exception being subject *D*. The two right-most sets of bars show the results for averaging over all individual models as well as from computing models from combining all the subjects' data, i.e. from building a general model instead of personal models.

Table 1 shows two example feature vectors, i.e. parts of two salience models, sorted by the values of the weights. These weights were obtained when training on all instances of subjects G and H, respectively. The different ordering of the features reflects different preferences of these two subjects when choosing landmarks. The highest type features of subject D are features associated primarily with streets, while for G, we can also find building features such as *church*, and entities of other types, e.g. *fountain*.

For all subjects, many type features are not contributing to the salience scores at all, they have a weight of 0. The number of type features that have non-zero weights ranges between 66 and 96. The positional and context features always have non-zero weights. The *distance* feature generally has a low weight value, meaning that closer landmarks will have higher salience scores. The *angle* features have weight values close to 0, both positive and negative. For the *duplicates* feature, the method generally learns weights around or below 0.

7. Discussion

The SVM Ranking method manages to mimic the user's salience preferences in 39% of the tested instances. How good is this result? Recall that we are aiming for an interactive guiding 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 recognise 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 four-fifths of the total number of segments). For example, subject G's models were successful in 60% of the test instances. On average, the Mean Reciprocal Rank is 0.77 for this subject. For subject H, where many more training instances were available, the method achieved a FHS rate of 39% and a Mean Reciprocal Rank of 0.60. 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' self-rated scores of overall familiarity with the area in Figure 3.

Table 1 shows parts of the 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 right-most part of Figure 3 shows the evaluation measures for training a model on four-fifths of all available data. We can see no improvement, which strengthens the plausibility of a personal salience model for each user.

Although the type features seem to differ in how they contribute to the salience scores, the *distance* feature shows a more clear tendency. All subjects are preferring landmarks close to their position. We expected the *duplicates* feature to have rather low weights, preferring objects that have unique type features. This expectation is generally met. That this weight is not as low as the lowest type features follows from the possible values for these features. Type features can only have a value of 1 or 0, while the *duplicates* feature can have any value from 0 to $C-1$ where C is the size of the candidate set.

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). A general model derived from previous users can be a good start. While our general model did not improve the individual models, it also is not considerably worse, ranking 40% of the landmarks correctly and scoring a mean reciprocal rank of well above 0.5.

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.

Notes

[1] We leave the incorporation of global landmarks for future research.
[2] http://www.speech.kth.se/hat/
[3] We are disregarding OSM *relations* for the time being.
[4] http://wiki.openstreetmap.org/wiki/Map_Features
[5] The page was accessed on Oct 5th 2015.
[6] The results for using the type features combined with the *duplicates* feature are similar to using the type features alone.

Funding

This work was supported by the Swedish Research Council [2013-4854].

ORCID

Jana Götze http://orcid.org/0000-0002-7829-5561
Johan Boye http://orcid.org/0000-0003-2600-7668

References

Allen, Gary L. 1997. "From Knowledge to Words to Wayfinding: Issues in the Production and Comprehension of Route Directions." In *Proceedings of the International Conference on Spatial*

Information Theory (COSIT), Vol. 1329 of Lecture Notes in Computer Science, edited by S. C. Hirtle and A. U. Frank, 363–372. Berlin Heidelberg: Springer.

Allen, Gary L. 2000. “Principles and Practices for Communicating Route Knowledge.” *Applied Cognitive Psychology* 14 (4): 333–359.

Boye, Johan, Morgan Fredriksson, Jana Götze, Joakim Gustafson, and Jürgen Königsmann. 2014. “Walk This Way: Spatial Grounding for City Exploration.” In *Natural Interaction with Robots, Knowbots and Smartphones*, edited by Mariani, Joseph, Rosset, Sophie, Garnier-Rizet, Martine, and Devillers, Laurence, 59–67. New York: Springer.

Bunescu, Razvan, and Marius Paşca. 2006. “Using Encyclopedic Knowledge for Named Entity Disambiguation.” In *Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, April 9–16.

Couclelis, Helen. 1996. “Verbal Directions for Way-finding: Space, Cognition, and Language.” In *The Construction of Cognitive Maps*, Vol. 32 of GeoJournal Library, 133–153. Netherlands: Springer.

Dahlbäck, Nils, and Arne Jönsson. 1989. “Empirical Studies Of Discourse Representations For Natural Language Interfaces.” In *Proceedings of the Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, edited by Harold L. Somers, and Mary McGee Wood, 291–298. Manchester: The Association for Computational Linguistics.

Daniel, M.-P., and Michel Denis. 1998. “Spatial Descriptions as Navigational Aids: A Cognitive Analysis of Route Directions.” *Kognitionswissenschaft* 7: 45–52.

Denis, Michel. 1997. “The Description of Routes: A Cognitive Approach to the Production of Spatial Discourse.” *Current Psychology of Cognition* 16 (4): 409–458.

Denis, Michel, Francesca Pazzaglia, Cesare Cornoldi, and Laura Bertolo. 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., G. V. Moon, T. Moore, and D. Jones. 1999. “Guiding Blind Pedestrians with a Personal Navigation System.” *Journal of Navigation* 52: 330–341.

Duckham, Matt, Stephan Winter, and Michelle Robinson. 2010. “Including Landmarks in Routing Instructions.” *Journal of Location Based Services* 4 (1): 28–52.

Elias, Birgit. 2003. "Extracting Landmarks with Data Mining Methods.” In *Proceedings of the International Conference on Spatial Information Theory (COSIT)*, Vol. 2825 of Lecture Notes in Computer Science, edited by W. Kuhn, M. F. Worboys, and S. Timpf, 375–389. Berlin Heidelberg: Springer.

Goetz, Marcus. 2012. “Using Crowdsourced Indoor Geodata for the Creation of a Three-dimensional Indoor Routing Web Application.” *Future Internet* 4 (2): 575–591.

“Google Maps Navigation. Google Inc.” 2013. http://www.google.com/mobile/navigation.

Götze, J., and J. Boye. 2013. “Deriving Salience Models from Human Route Directions.” In *Workshop on Computational Models of Spatial Language Interpretation and Generation 2013 (CoSLI)*, 36–41.

Götze, J., and J. Boye. 2015a. “Resolving Spatial References using Crowdsourced Geographical Data.” In *Proceedings of the 20th Nordic Conference of Computational Linguistics (NODALIDA)*, May, 61–68. Linköping University Electronic Press.

Götze, Boye, and Johan Boye. 2015b. “Turn Left” Versus “Walk Towards the Café”: When Relative Directions Work Better Than Landmarks.” In *AGILE 2015*. Lecture Notes in Geoinformation and Cartography, edited by Fernando, Bacao, Maribe, l Yasmina Santos, and Marco, Painho. 253–267. Springer International Publishing. doi: 10.1007/978-3-319-16787-9_15.

Haklay, M., and P. Weber. 2008. “OpenStreetMap: User-generated Street Maps.” *Pervasive Computing, IEEE* 7 (4): 12–18.

Helal, A., S. E. Moore, and B. Ramachandran. 2001. “Drishti: An Integrated Navigation System for Visually Impaired and Disabled.” In *Proceedings of the Fifth International Symposium on Wearable Computers*, 149–156.

Hentschel, Matthias, and Wagner, Bernardo. 2010. “Autonomous Robot Navigation Based on OpenStreetMap geodata.” In *Proceedings of the 13th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, September 1645–1650.

Hill, Robin, Jana Götze, and Bonnie Webber. 2012. *Final data release, Wizard-of-Oz (WoZ) experiments*, http://www.spacebook-project.eu/pubs/D6.1.2.pdf.

Janarthanam, Srinivasan, Oliver Lemon, Xingkun Liu, Phil Bartie, William Mackaness, Tiphaine Dalmas, and Jane Goetze. 2012. “Integrating location, visibility, and Question-Answering in a spoken dialogue system for pedestrian city exploration.” In *Proceedings of the 16th Workshop on the Semantics and Pragmatics of Dialogue (Semdial)*.

Joachims, Thorsten. 2002. “Optimizing Search Engines Using Clickthrough Data.” In *Proceedings of the ACM Conference on Knowledge Discovery and Data Mining (KDD)*, 133–142. Finland: Helsinki.

Joachims, Thorsten. 2006. “Training Linear SVMs in Linear Time.” In *Proceedings of the ACM Conference on Knowledge Discovery and Data Mining (KDD)*, edited by Tina Eliassi-Rad, Lyle H. Ungar, Mark Craven, and Dimitrios Gunopulos, 217–226.

Kennedy, Alistair, and Diana Inkpen. 2006. “Sentiment Classification of Movie Reviews Using Contextual Valence Shifters.” *Computational Intelligence* 22 (2): 110–125.

Kennedy, Alan, Alan Wilkes, Leona Elder, and Wayne S. Murray. 1988. “Dialogue with machines.” *Cognition* 30 (1): 37–72.

Lovelace, Kristin L., Mary Hegarty, and Daniel R. Montello. 1999. “Elements of Good Route Directions in Familiar and Unfamiliar Environments.” In *Spatial Information Theory. Cognitive and Computational Foundations of Geographic Information Science*, Vol. 1661 of Lecture Notes in Computer Science, edited by Christian Freksa, and David M. Mark, 65–82. Berlin Heidelberg: Springer.

Lynch, Kevin. 1960. *The Image of the City*. Cambridge, MA: MIT Press.

Mast, Vivien, Jan Smeddinck, Anna Strotseva, and Thora Tenbrink. 2010. “The Impact of Dimensionality on Natural Language Route Directions in Unconstrained Dialogue.” In *Proceedings of the 11th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGdial)*, 99–102. Stroudsburg, PA, USA, Tokyo, Japan: Association for Computational Linguistics.

May, Andrew J., Tracy Ross, Steven H. Bayer, and Mikko Tarkiainen. 2003. “Pedestrian Navigation AIDS: Information Requirements and Design Implications.” *Personal and Ubiquitous Computing* 7 (6): 331–338.

Michon, Pierre-Emmanuel, and Michel Denis. 2001. “When and Why Are Visual Landmarks Used in Giving Directions?” In *Spatial Information Theory, Vol. 2205 of Lecture Notes in Computer Science*, edited by Daniel R. Montello, 292–305. Berlin Heidelberg: Springer.

Nothegger, Clemens, Stephan Winter, and Martin Raubal. 2004. “Selection of Salient Features for Route Directions.” *Spatial Cognition & Computation* 4 (2): 113–136.

Presson, Clark C., and Daniel R. Montello. 1988. “Points of Reference in Spatial Cognition: Stalking the Elusive Landmark.” *British Journal of Developmental Psychology* 6: 378–381.

Radev, Dragomir R., Hong Qi, Harris Wu, and Weiguo Fan. 2002. “Evaluating Web-based Question Answering Systems.” In *Proceedings of the International Conference on Language Resources and Evaluation (LREC)*. European Language Resources Association.

Raubal, Martin, and Stephan Winter. 2002. “Enriching Wayfinding Instructions with Local Landmarks.” In *Geographic Information Science*, Vol. 2478 of Lecture Notes in Computer Science, edited by Max J. Egenhofer, and David M. Mark, 243–259. Berlin Heidelberg: Springer.

Rehrl, Karl, Elisabeth Häusler, and Sven Leitinger. 2010. “GPS-Based Voice Guidance as Navigation Support for Pedestrians, Alpine Skiers and Alpine Tourers.” In *Proceedings of Workshop on Multimodal Location Based Techniques for Extreme Navigation*, 13–16.

Rehrl, Karl, Sven Leitinger, Georg Gartner, and Felix Ortag. 2009. “An Analysis of Direction and Motion Concepts in Verbal Descriptions of Route Choices.” In *In Spatial Information Theory*, Vol. 5756 of Lecture Notes in Computer Science, edited by Hornsby, Kathleen Stewart, Claramunt, Christophe, Denis, Michel, and Ligozat, Gérard, 471–488. Berlin Heidelberg: Springer.

Ross, Tracy, Andrew J. May, and Simon Thompson. 2004. “The Use of Landmarks in Pedestrian Navigation Instructions and the Effects of Context.” In *Mobile HCI*, Vol. 3160 of Lecture Notes in Computer Science, edited by Stephen. A. Brewster, and Mark D. Dunlop, 300–304. Springer.

Schroder, Catherine J., William A. Mackaness, and Bruce M. Gittings. 2011. “Giving the ‘Right’ Route Directions: The Requirements for Pedestrian Navigation Systems.” *Transactions in GIS* 15 (3): 419–438.

Sorrows, Molly E., and Stephen C. Hirtle. 1999. “The Nature of Landmarks for Real and Electronic Spaces.” In *Proceedings of the International Conference on Spatial Information Theory (COSIT)*, Vol.

1661 of Lecture Notes in Computer Science, edited by Freksa, Christian, and David M. Mark, 37–50. Berlin Heidelberg: Springer.

Tom, Ariane, and Michel Denis. 2004. "Language and Spatial Cognition: Comparing the Roles of Landmarks and Street Names in Route Instructions." *Applied Cognitive Psychology* 18 (9): 1213–1230.

Tom, Ariane Cecile, and Barbara Tversky. 2012. "Remembering Routes: Streets and Landmarks." *Applied Cognitive Psychology* 26 (2): 182–193.

Wolfensberger, Marius, and Kai-Florian Richter. "A Mobile Application for a User-Generated Collection of Landmarks." In *Web and Wireless Geographical Information Systems*, Vol. 9080 of Lecture Notes in Computer Science, edited by Gensel, Jérôme, and Tomko, Martin, 3–19. Cham: Springer International Publishing.