A Probabilistic Approach to Modelling Spatial Language with Its Application To Sensor Models

Jamie Frost[1], Alastair Harrison[2], Stephen Pulman[1], and Paul Newman[2]

[1] University of Oxford, Computational Linguistics Group, OX1 3QD, UK
{jamie.frost,sgp}@clg.ox.ac.uk
[2] University of Oxford, Mobile Robots Group, OX1 3PJ, UK
{arh,pnewman}@robots.ox.ac.uk

Abstract. We examine why a probabilistic approach to modelling the various components of spatial language is the most practical for spatial algorithms in which they can be employed, and examine such models for prepositions such as 'between' and 'by'. We provide an example of such a probabilistic treatment by exploring a novel application of spatial models to the induction of the occupancy of an object in space given a description about it.

1 Introduction

Space occupies a privileged place in language and our cognitive systems, given the necessity to conceptualise various semantic domains. Spatial language can broadly be divided into two categories [1]: functions which map regions to some part of it, e.g. 'the corner of the park', and functions (in the form of spatial prepositions) which map a region to either an adjacent region, projection or axis, e.g. 'the car between the two trees'. Approaches to implementing spatial models have fallen into two categories. [2] for example takes a logic-based approach, using a set of predicates on objects and binary or tertiary relations that connect objects to generate descriptions of objects that distinguishes it from others. A second approach is a numerical one, which given some reference object or objects and another 'located' object[1] or point, assigns a value based on some notion of 'satisfaction' of the spatial relation in question. But conceptualisation of this assigned value has a large amount of variety. [3] uses a 'Potential Field Model' characterised by potential fields which decreases away from object boundaries. [4] for example uses a linear function to model topological prepositions such as 'near', and produces a value in the range [0,1] depending on whether some point is directly by the object in question or on/beyond the horizon.

However, we argue that a conceptually more rigorous probabilistic approach is needed for all aspects of spatial language, in which validity of some spatial or semantic proposition is determined by the likelihood a human within the context

[1] We use the term 'locative expression' to refer to any expression whose intention is to identify the location of an object or objects (such as 'a chair by the table'). The 'located object' refers to the object in question, and the 'reference' object(s) are others that can be used to determine the location of the located object (the *table* in the latter example).

of the expression would deem it to be true. We motivate this by the following reasons:

1. It provides a uniform treatment of confidence across both spatial and non-spatial domains; uncertainty may be established in the latter in cases of variants of descriptive attributes (such as names) for example. As a result these models can be used in a variety of spatial algorithms such as searching or describing objects and inferring the occupancy in space of an object.
2. In the latter of the above applications (which will be explored in detail) as well as other independent systems or frameworks, a probabilistic representation is often required.
3. Combining multiple spatial observations becomes more transparent: While any monotonically increasing or decreasing function is sufficient to establish a relative measure of applicability across candidate points or objects, the lack of consideration of the function's 'absolute' value becomes problematic when combining data from different spatial models, for example if we were to say 'The chair is by the table *and* between the cat and the rug'.

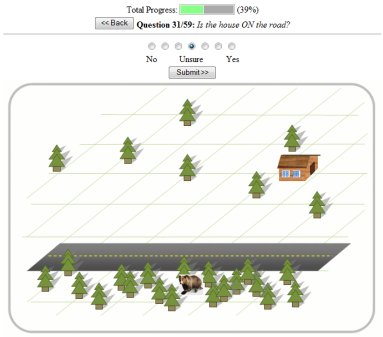

Fig. 1. One of the questions for the 'on/next to/by' section in an online experiment. There were 132 questions in total across the 3 sections.

Such an approach of assessing the 'acceptibility' of regions given a spatial relation is based on a concept called 'Spatial Templates' established by [5], but a probabilistic approach puts more emphasis on *absolute* value. What precisely then do we mean by 'human confidence'? One might think we can measure it by the probability that a given human would consider a (spatial) proposition to be true. But such a notion neglects a concept in philosophy known as *subjectivism*, in which rational agents can have degrees-of-belief in a proposition (rather than constricted to boolean answers of 'agree' or 'disagree'), and probabilities can be interpreted as the measure of such a belief. With such an assumption it is therefore sufficient to construct our models based on the 'average degree-of-belief' across people in some sample. Generically, this confidence can be defined as $p(\phi|\psi)$, where ϕ represents the proposition and ψ represents the context. For a particular spatial model, one might use $p(in\_front(obj_1, obj_2)|x_t)$, where x_t is the current position of the observer. We use $\phi_{\mathbf{x}}$ as a convenience to indicate that the location (say its centre of mass) of the *located object* in ϕ is at position $\mathbf{x}$.

In the next section we present such models we have developed for the prepositions 'between' and 'by', and present a possible novel approach in which we

might induce the occupancy of an object in space given a spatial description.We carried out an online experiment in which users asserted the validity of various locative expressions given a variety of scenes. For each category of spatial relation, e.g. *by* and *between* (and a number of other prepositions not presented here), the user was asked to rate the extent to which they agreed with the given statement, on a scale of 1 (representing 'no') to 7 (representing 'yes'), each question accompanied by a picture[2]. To produce the 'average degree-of-belief' we

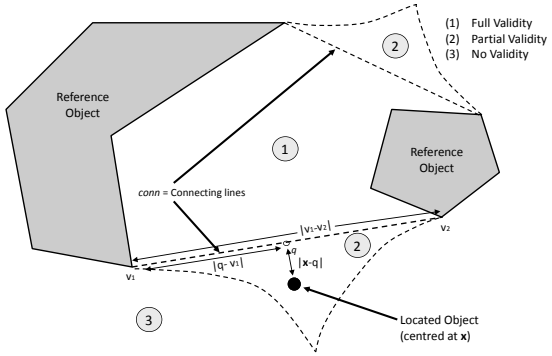

Fig. 2. The variation of confidence for the preposition 'between'.

scaled the average answer to [0,1]. Our models are based on the *Proximal Model* as described in [7]. That is, features are based on the nearest point to the reference object, thus incorporating the shape of the object. This is in contrast to the *Centre-of-Mass Model* (as used in [4] for example) which treats all objects as points. This latter approach is computationally simpler and requires less data, although can be problematic for larger objects; if for example we were to assess the acceptibility of 'you are near the park', we would expect such a judgement to be based on proximity to the edge of the park rather than the centre.

2 Spatial Models for 'between' and 'by'

2.1 "Between"

The model we present below determines the acceptibility of a proposition $\phi = between(a, b, c)$ such that a is the located object, b and c the reference objects, and the position of a is at $\mathbf{x}$. We determined that any point within the convex hull of the two reference objects (excluding the area of the objects themselves) was deemed to be fully valid. Outside of this area, certainty degraded proportional to the centrality of the object. Our model below quantifies these findings:

$$p(\phi_{\mathbf{x}}|x_t) = p(\phi_{\mathbf{x}}) = \begin{cases} max(0, 1 - \frac{|\mathbf{x}-q|}{tol}) \text{ if } \mathbf{x} \notin Hull(ref_1 \cup ref_2) \\ 1 \text{ otherwise} \end{cases} \quad (1)$$

[2] The experiment was restricted to native English speakers only, due to cross-linguistic variations in spatial coding, such as a lack of distinction between different frames of reference (that is, distinguishing between say the deictic interpretation of "in front of the tree" based on the position of the observer, and the intrinsic interpretation based on the salient side of an object, as in "in front of the shop") [6].

$$\text{s.t.} \quad q = \arg\min_{q'}\{|\mathbf{x} - q'||q' \text{ on line } l\}, \quad tol = |v_1 - v_2|k_1(\frac{|q - v_1|}{|v_1 - v_2|})^{k_2}$$

$$l : (v_1, v_2) = \arg\min_{l'}\{|\mathbf{x} - q'||q' \text{ on line } l', l' \in conn\}$$

$$conn = \{(\bar{v}_1, \bar{v}_2)|\bar{v}_1 \in ref_1, \bar{v}_2 \in ref_2, (\bar{v}_1, \bar{v}_2) \in edges(Hull(ref_1 \cup ref_2))\}$$

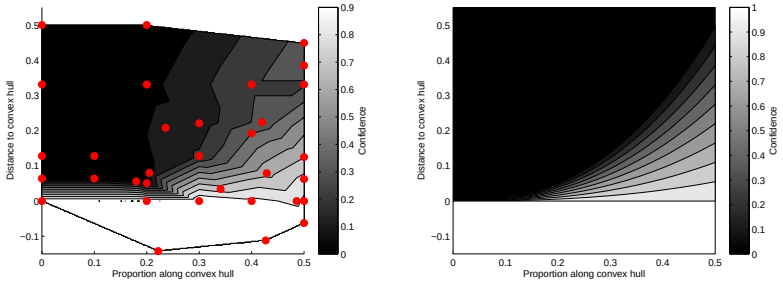

Fig. 3. A comparison of experimental results against the inferred model for 'between'. Both the x and y axis are in terms of the length of the convex hull edge l.

$\mathbf{x}$ is the central point of the located object in question, ref_1 and ref_2 are the vertices of the two referenced objects, $Hull(V)$ gives the convex hull of the set of vertices V (thus q is the nearest point on the convex hull to $\mathbf{x}$), tol gives the maximum allowed distance from the convex hull before the confidence score is 0, $conn$ is the set of 2 edges on the convex hull which connect the shapes corresponding to ref_1 and ref_2 (that is, the straight dotted lines in Fig. 2 and function $edges$ gives the edges of a polygon. k_1 controls the maximum tolerance permitted, a specified proportion of the distance between the two objects, and k_2 controls the curvature of this ambiguous region. Via model fitting (using the minimum sum of squared differences) we found values of $k_1 = 0.55$ and $k_2 = 2.5$ yielded the best results (see Figure 3).

2.2 "By"

For the preposition 'by', there are 3 main variables that can influence the magnitude of the confidence score; the base width (w) and height (h) of the reference object, and the distance (d) from the reference object. For polygonal objects, users were given 8 different reference objects in their scenarios, of a variety of different widths and heights. It was found that although the confidence score for a given distance with respect to the width of the object (i.e. $\frac{d}{w}$) was a good starting point (see Fig 4(a)), greater heights led to a small increase in probability. Assuming a linear relationship with height (again relative to the object width), we therefore divide by $\frac{h}{w} + k_h$ for some constant k_h (given that flat objects such as lakes still yield a non-zero confidence score). Additionally, smaller objects tended to have a larger tolerance of distance with respect to this width, although this effect became less prominent as the width of the object became

very large. Thus we multiply the distance by $\log(w + k_w)$ for some constant $k_w \geq 1$, since for very small objects we still expect some tolerance of distance. Combining these relationships and simplifying, we suggest the following model:

$$p(\phi_{i,j}|x_t) = clamp(k_c - k_m d \frac{\log(w + k_w)}{h + k_h w}) \quad (2)$$

where k_m and k_c the coefficients of some line to obtain the confidence from the adjusted distance, and *clamp* clamps the overall value to the range $[0, 1]$. Fig. 4(b) shows the effect of these using these transformations, using $k_w = 14$ and $k_h = 2$, resulting in values for k_m and k_c of 1.38 and 1.15 respectively. Ultimately it is impossible to base any model of 'by' on physical metrics alone; the 'use case' of objects, i.e. the set of contexts in which an object is used, is likely to have an effect. In Fig. 4(b) for the case where the reference object was a chair, it is apparent confidence deteriorated with distance much faster than expected. But if one considers that a chair is intended 'to be sat on', and therefore adjust the recorded height h to the more salient 'seat-level', we obtain confidence values very close to the model for this example.

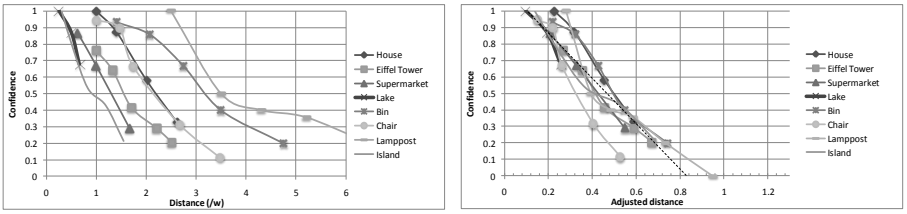

Fig. 4. Experimental results for the preposition 'by' for a number of different examples. Each series indicates the *reference object* used in the locative expression, e.g. 'lake' in "The house is by the lake". Graph (a) shows distance (in terms of width) plotted against validity. Graph (b) shows adjustments as per equation 2.

3 Occupancy Grid Maps

We now propose a method to infer the occupancy in space of a particular object given an observation in the form of a spatial description made about it. This is strongly predicated on a probabilistic treatment of our spatial models discussed earlier. Occupancy Grid Mapping is a technique employed in robotics to generate maps of an environment via noisy sensor measurements. The occupancy grid map is useful because it can subsequently be fused with other maps obtained from say physical sensors. The aim is to produce a posterior $p(m|z_{1:t}, x_{1:t})$ where $m = \{m_i\}$ is a partitioning of space into a finite grid of cells m_i, $z_{1:t}$ are the observations made up to time t, and $x_{1:t}$ are the poses of the robot at each observation. m_i is the event that cell i is occupied, thus $p(m_i)$ describes the probability that cell i is occupied. In the scope of this paper, we focus on how the 'inverse sensor model' $p(m_i|z_t, x_t)$ can be computed, although a more detailed description of Occupancy Grid Mapping can be found in [8].

Our aim is to compute this inverse sensor model, in terms of our calculated $p(\phi|x_t)$ probabilities from the previous section. An important simplifying assumption we make is that locative expressions refer to a *specific point*

in space, within the boundaries of the object in question. This seems intuitive; were we to describe a town as being '10km away', it would clearly be fallacious to assume that the entirety of the town is precisely 10km away. We define a probability $p(r_{i,j}|x_t, z_t)$, where $r_{i,j}$ represents the event that the observer was referring to a point (i, j) in their observation, and z_t is the locative expression such that the position of the located object is not specified, say $\phi_\ominus$ (since we consider such a position in $r_{i,j}$). We can then calculate the desired probability easily by simply normalising our confidence function across the space: $p(r_{i,j}|x_t, z_t = \phi_\ominus) = \alpha p(\phi_{i,j}|x_t)$. Before we determine how to calculate $p(m_{i,j}|x_t, z_t)$, we analyse the conceptual parallelism between traditional Occupancy Grid Mapping and that employed in our linguistic context.

3.1 A Comparison of Sensor Models

On a cursory inspection there are some initial clear similarities that can be drawn between the traditional occupancy grid map and our linguistic variant. Both involve the pose of some observer x_t (although depending on the spatial model this is sometimes irrelevant) and some manifestation of an observation z_t; a physical sensor reading with respect to the traditional approach and a locative expression for the linguistic approach. Upon closer analysis more similarities can be drawn. With a physical sensor, we expect a measurement of distance to a point being sensed to be noisy, and thus maintain a probability distribution with regards to the precise position of this point. This corresponds to our distribution $p(r_{i,j}|x_t, z_t)$. For a locative expression of a town being "10km away", human error or rounding is likely to lead to uncertainty in the judged distance, and additionally the direction of the town is unspecified, leading to a 'blurred doughnut' type distribution.

There are however a number of conceptual differences. With traditional Occupancy Grid Mapping the posterior for a cell is only updated if it was part of the sensor range (i.e. we make no assumption with regards to space outside the limited range of our sensor). With locative expressions however, we can infer data outside that explicitly conveyed. Suppose for our town example, the town was 1km in diameter, and that the distance judgement of 10km (to some point within the town) was entirely accurate. If the centre of the town was actually 10.5km away, our observation would still hold, but a point any further could not possibly be occupied by 'town'.

3.2 Computing the Inverse Sensor Model

We can use the above fact to compute $p(m_{i,j}|x_t, z_t)$ from our previously calculated values of $p(r_{i,j}|x_t, z_t)$. Let $\mathcal{Q}$ be the set of possible 'poses' for the located object such that the point (i, j) is within the object's boundary, and a pose is the position and orientation of the object. Given our assumption that the observer referred to a point within the confines of their perceived position of the object, $\mathcal{Q}$ represents all valid poses of the object given such a point. It follows that $p(m_{i,j}|x_t, z_t) = \int_{q\in\mathcal{Q}} p(q|x_t, z_t)\,\mathrm{d}q$. For each pose $q \in \mathcal{Q}$ there is an associated frame (i^*_q, j^*_q, θ_q) where (i^*_q, j^*_q) is the *nominal centre* of the shape in pose q (say

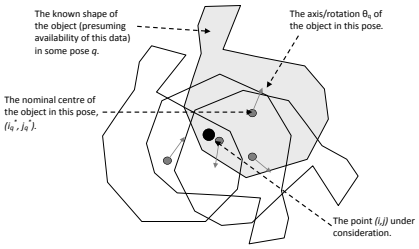

Fig. 5. In calculating the occupancy probability $p(m_{i,j}|x_t, z_t)$, we consider all possible poses of the located object in which the point (i, j) is confined. The probability of each pose q is $p(r_{i_q^*,j_q^*}|x_t, z_t)p(\theta_q)$.

the centre of mass) and θ_q is the rotation of the object about this point. It is then possible to use $p(r_{i_q^*,j_q^*}|x_t, z_t)$ to refer to the probability of the object being positioned at (i_q^*, j_q^*) (see Fig. 5). The pose also has a probability $p(\theta_q)$ associated with its orientation; for simplicity we assume this is independent of x_t and z_t (although the use of $p(\theta_q|z_t)$ would allow us to model for example observations such as "The boat is in front of you, *facing East*"). Putting this together, this gives us the following equation to compute the occupancy probability:

$$p(m_{i,j}|x_t, z_t) = \int_{q\in\mathcal{Q}} p(r_{i_q^*,j_q^*}|x_t, z_t)p(\theta_q)\,\mathrm{d}q \quad \text{s.t. } \mathcal{Q} = \{q \,|\, (i,j) \in R(q)\} \tag{3}$$

where $R(q)$ is the region of the located object in pose q. Considering the pose of the located object has useful consequences; it allows us to model for example that vehicles are aligned to the direction of a road. Given a lack of prior shape information with regards to the located object, and given the above integral is somewhat intractable, a suitable approximation is to use the approximate width W of the object (which can be obtained via knowledge of the class of the located object, say the usual width of a town). If we infer as little about the shape as possible, the resulting approximation of the shape is a circle of diameter W. Equation 3 then reduces to the following:

$$p(m_{i,j}|x_t, z_t) = \int_{(i',j')\in R(\frac{W}{2},i,j)} p(r_{i',j'}|x_t, z_t)\,\mathrm{d}i'\,\mathrm{d}j' \tag{4}$$

s.t. $R(\frac{W}{2}, i, j)$ is a set of points in a circular region of centre (i, j) and radius $\frac{W}{2}$

4 Conclusions & Future Work

In this paper we motivated a probabilistic approach to modelling spatial language that can be used in a number of algorithms, and provided an example of such an algorithm to induce a sense of 'the space that an object occupies' via the use of occupancy grid maps. We also presented models for the prepositions

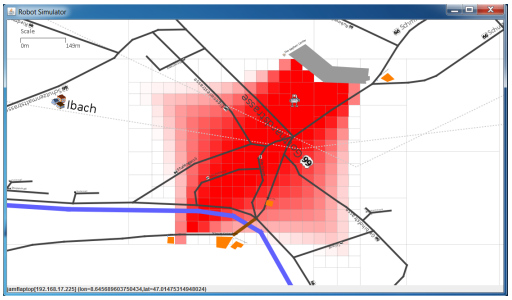

Fig. 6. The occupancy grid generated by a 'between' observation for two objects in an environment. Note that the grid consists of cells of variable size; such a modification to the OGM allows us to choose a cell size appropriate to the scale of the observation, as well as represent areas of constant probability or empty space efficiently.

'by' and 'between' based on the results of an online experiment. Future work is predominantly focused further development of our dialogue manager language that interacts with these spatial models, as well as developing further algorithms which make use of such models. For example, we developed an algorithm that combines semantic and spatial models to provide confidence scores for arbitrarily complex locative expressions (including those based on current bounded trajectories, such as 'the second left'). We are also investigating a measure of 'relevance' (one of the Gricean maxims [9]) for locative expressions, a consideration that is particularly key in generating descriptions of objects or locations.

This work has been supported by the European Commission under grant agreement number FP7-231888-EUROPA.

References

1. Herskovits, A. (1986) Language and Spatial Cognition, Cambridge University Press, .
2. Dale, R. and Haddock, N. (1991) In Proceedings of the fifth conference on European chapter of the Association for Computational Linguistics Morristown, NJ, USA: Association for Computational Linguistics. pp. 161–166.
3. Olivier, P. and Tsujii, J.-I. (2004) *Artificial Intelligence Review* **8**, 147–158.
4. Kelleher, J. D. and Costello, F. J. (2009) *Comput. Linguist.* **35(2)**, 271–306.
5. Logan, G. and Sadler, D. Language and space chapter A computational analysis of the apprehension of spatial relations, pp. 493–529 MIT Press (1996).
6. Levinson, S. C. and Wilkins, D. P. Grammars of Space: Explorations in Cognitive Diversity chapter 1, pp. 4–5 Cambridge University Press (2006).
7. Regier, T. and Carlson, L. A. (2001) *J. Exp Psychol Gen* **130(2)**, 273–298.
8. Thrun, S. (2003) *Auton. Robots* **15(2)**, 111–127.
9. Grice, H. P. (1975) Logic and conversation In Peter Cole and Jerry L. Morgan, (ed.), Syntax and semantics 3: Speech Acts, volume **3**, pp. 41–58 New York: Academic Press.