Fusion of Natural Language Propositions: Bayesian Random Set Framework

Adrian N. Bishop  
NICTA, Canberra Research Lab  
Australian National University (ANU)  
Canberra, Australia  
Email: adrian.bishop@nicta.com.au

Branko Ristic  
ISR Division  
Defence Science and Technology Organisation (DSTO)  
Melbourne, Australia  
Email: branko.ristic@dsto.defence.gov.au

Abstract—This work concerns an automatic information fusion scheme for state estimation where the inputs (or measurements) that are used to reduce the uncertainty in the state of a subject are in the form of natural language propositions. In particular, we consider spatially referring expressions concerning the spatial location (or state value) of certain subjects of interest with respect to known anchors in a given state space. The probabilistic framework of random-set-based estimation is used as the underlying mathematical formalism for this work. Each statement is used to generate a generalized likelihood function over the state space. A recursive Bayesian filter is outlined that takes, as input, a sequence of generalized likelihood functions generated by multiple statements. The idea is then to recursively build a map, e.g., a posterior density map, over the state space that can be used to infer the subject state.

Keywords: Spatial prepositions, natural language, information fusion, Bayesian estimation, random set theory.

I. INTRODUCTION

Natural language processing involves the design of algorithms for understanding and processing human, naturally conveyed, statements and prepositions1 [1]. Note that such processing typically goes beyond simple speech, or text recognition and involves the interpretation of natural language for decision and control sequences.

This work concerns an automatic information fusion scheme for state estimation where the inputs (or measurements), used to reduce the uncertainty in the state of a subject, are in the form of natural language propositions. The following example is indicative of the scenario motivating this work.

Example 1. Imagine a battlefield scenario in which a number of squads are scattered about the field. Suppose there is an enemy mortar in the field whose position is unknown but observed by three squad leaders without accurate positioning equipment. The platoon commander instructs the squad leaders to relay the location of the mortar as they observe it. Consider the three statements

* the mortar is behind the stone wall
* from my position the mortar is next to the barn
* I think the mortar is near the front of the barn spoken by the first, second and third squad leader respectively.

Now, given a map of the battlefield that contains the required landmarks, i.e. the stone wall and barn, along with the relevant positions of the squad leaders, then it follows that the platoon leader can infer the enemy mortar location. This work looks at the design of automatic systems to combine statements such as those in this example into a unified spatial representation over the space of interest on which one can infer the state (e.g. location, velocity) of certain subjects.

In particular, we consider spatially referring expressions concerning the spatial location (or state value) of certain subjects of interest with respect to known anchors in a given state space; an anchor is a subject with a fixed and known state value. The space of interest in the preceding example was the battlefield, e.g. a subset of $\mathbb{R}^2$, but more abstract spaces and problems fit within the proposed framework. Each statement leads to a generalized likelihood function on the state space. The idea is then to use such likelihood functions to recursively build a map, e.g., a posterior density map, over the state space that can be used to infer the subject state.

Natural language statements and, in particular, spatial prepositions are typically ambiguous and depend greatly on the context and grounding of the subjects referenced [2]–[4]. For example, if we say, “The ball is in front of the car”, it can mean that we want to locate the ball in relation to the car from the point of view of the speaker, with respect to the orientation of the car itself, or with respect to the actual direction of the motion of the car [2]. In addition to the various hypotheses concerning the use of “in front” there is also uncertainty in each hypothesis in that the relationship “in front” is geometrically dependent on the configuration of the speaker, car and even the listener. For example, if the speaker is close to the car then the ball should be closer to the car than if the speaker is further away. Each hypothesis should have a smaller variance in this case.

The result is that any likelihood function for the state of the subject, e.g. the ball in the previous case, that takes such a preposition as input must be multi-modal to account for the multiple hypotheses in the interpretation. The likelihood function must also allow for the uncertainty in the geometrical nature of the spatial relation itself with respect to each hypothesis [4]–[6].

1We also use the notion of a proposition, as opposed to a preposition, to refer to particular declarative sentences etc with a corresponding truth value.
There has been some work in the linguistic and robotics communities that turns natural language statements, or spatially referring expressions, into spatial representations suitable for, e.g., inference on spatial relationships between objects and human-robot interaction [7]–[10]. Our work differs from this existing work in that we seek to develop a rigorous probabilistic framework in which one can form a mathematically general likelihood function from certain natural language statements and then perform recursive Bayesian filtering and inference. Our work is motivated by the discussions and mathematical formalism introduced in [11], [12]. In particular, we employ the random-set formulation of [11], [12] to form the generalized likelihood functions in this work. The Bayesian fusion algorithm outlined in this work subsumes as a special case, Dempster-Shafer theory, fuzzy set theory, Bayesian fusion with likelihood mixture models etc [11]–[16].

II. MODELLING THE POSITION OF SUBJECTS IN SPACE

Fix an underlying Borel measurable space \((S, B(S))\) where \(B(\cdot)\) is a Borel \(\sigma\)-algebra [17]. The space \(S\) is the state space. A subject of interest is denoted by \(\mathcal{S} = \{\text{target}\}\) and \(\Sigma \in \mathbb{R}^3\).

Let \(\{\phi_i\}_{i=1}^n\) denote a set of propositions concerning the value of the state of a subject. To avoid being tied to a particular semantic representation we take the simplistic view in this work to model \(\phi\) by

\[
\phi^* = \text{the subject is located with some spatial relationship to an anchor in the space} \quad (1)
\]

Any \(\phi\) which is homomorphic to such a form is acceptable\(^2\). Thus, we refer to \(\phi^*\) as the normal form of the proposition. We write \(\phi_i \sim \phi^*\) if \(\phi_i\) is in a normal form. Associated with each proposition \(\phi_i\) is a map \(\varphi_i : \phi_i \to [0, 1]\) resulting in the tuple \((\phi_i, \varphi_i)\). The map \(\varphi_i\) is like a probabilistic confidence or truth value of the proposition. If \(\varphi_i = 0\) then the proposition \(\phi_i\) can typically be neglected.

We consider a set of spatial relationships denoted by \(\{R_j\}_{j=1}^3\) and a set of anchors \(\{A_k\}_{k=1}^a\) with known positions \(\{a_k\}_{k=1}^a \subseteq S \times \mathbb{R}^3\) and \(A_i \neq A_j, \forall i \neq j\). Here \(\mathbb{R}^3\) may be a null-space (but \(S^1\) is a typical non-null example).

The universe of all spatial relationships and anchors is \(\mathcal{R}\) and \(\mathcal{A}\). Given a proposition \(\phi_i\) then the operator notation \(R(\phi_i)\) and \(A(\phi_i)\) respectively pulls out the spatial relationship and anchor referenced in \(\phi_i\).

A more specific example is then:

\[
\phi_i = \text{the error in the feedback loop is near the origin} \quad (2)
\]

where \(\mathcal{S} = \{\text{error}\}\), the spatial relationship is \(\mathcal{R} = \{\text{near}\}\) and \(\mathcal{A} = \{\text{origin}\}\) is the anchor. Another example is:

\[
\phi_i = \text{the state is on the boundary of the manifold} \quad (3)
\]

\(^2\)A discussion on the linguistic justification for such an approach is provided in the appendix.

Note that the set of anchors are subjects in \(\mathcal{S}\) with known states. Propositions of the form \(\phi_i\) are not spoken in isolation. They are spoken by an individual, the speaker, in a state \(s_i \in S \times \mathcal{B}(s)\) to another individual, the listener\(^3\), in state \(p_i \in S \times \mathcal{C}\). Thus, intrinsically associated with each proposition \(\phi_i\) is a state \(s_i \in S \times \mathcal{B}(s)\) and a state \(p_i \in S \times \mathcal{C}\). Both \(\mathcal{B}(s)\) or \(\mathcal{C}\) may be null-spaces (but also \(S^1\) are a typical non-null example spaces). The set of anchors is typically augmented with the state of the speaker and the listener.

We will make the following standing assumption.

Assumption 1. Each \(\phi_i\) is in the present tense. The state of the anchors are known. The subject and anchor are referred to singularly. The state \(s_i\) and state \(p_i\) are known.

We will neglect the problem of reference resolution in this work [9], [10]. Each proposition \(\phi_i\) leads to likelihood function on \(\mathcal{S}\) determined by the particular spatial relationship \(R_j\), the state of the anchors, the speaker and the listener. In this work we model such a likelihood via a sum of the form

\[
g(\phi_i|\Sigma) \equiv (1 - \varphi_i) + \sum_{k_i \in H_i} w_{k_i} \gamma_{k_i}(\phi_i, \cdot) \quad (4)
\]

where \(\sum_{k_i} w_{k_i} = \varphi_i\) and \(H_i\) is a function of \(R_j\) and \(A_k\). The functions \(\gamma_{k_i}(\phi_i, \cdot)\) are dependent on the parameters defining \(\phi_i\) and possibly some additional tuning terms. More will be said about how we construct \(g(\phi_i|\Sigma)\) later. However, as an example \(\gamma_{k_i}(\phi_i, \cdot)\) may be a probability density and then \(g(\phi_i|\Sigma)\) is a mixture density model.

Example 2. The situation is best described by an example:

\[
\phi_i = \text{the target is in front of the red car} \quad (5)
\]

where \(\mathcal{S} = \{\text{target}\}\), the spatial relationship is \(\mathcal{R} = \{\text{in front}\}\) and \(\mathcal{A} = \{\text{red car}\}\) is the anchor. Also \(\varphi_i = 1\).

Let \(\Sigma = \mathbb{R}^2\). Suppose the location of the car \(A \in \mathbb{R}^2 \times S^1\) has a position in \(\mathbb{R}^2\) and an orientation. For this example, take \(A = [x\ y\ \theta]^\top = [0 \ 0 \ 0]^\top\), such that the car is facing toward the positive \(x\)-axis. The speaker is at \(s_i = [0 \ 2.5\ 0]^\top\) and the listener is at \(p_i \in S\), i.e. for this example we won’t care where the listener is. The likelihood function is a sum of two Gaussian density functions

\[
g(\phi_i|\Sigma) \equiv \frac{1}{2} \gamma_1(\Sigma - q_{\text{1}}, \Omega_1) + \frac{1}{2} \gamma_2(\Sigma - q_{\text{2}}, \Omega_2) \quad (6)
\]

where

\[
\gamma(x - \mu, \Xi) = \frac{1}{(2\pi)^{d/2} |\Xi|^{1/2}} \exp\left(\frac{1}{2} \| \Xi^{-1/2} (x - \mu) \|_2^2 \right) \quad (7)
\]

and \(\Xi\) is the variance and \(\mu\) is a mean. The means \(q_{\text{1}}, q_{\text{2}}\) and variances \(\Omega_{\text{1,2}}, k_i \in H_i\) are defined based on \(\phi_i\). One example likelihood function is shown in Figure 1.

It follows that the proposition \(\phi_i\) results in two hypotheses regarding the position of the car generated by the spatial

\(^3\)Of course, there may exist multiple listeners but for simplicity we assume the speaker is speaking to one particular individual.
relationship \( R = \{\text{in front}\} \). The target may be in front of the car proper or in front of the car with respect to the speaker\(^4\). The position and variance of the Gaussian components in this example are chosen for illustrative purposes.

It follows that each \( k_i \in \mathcal{H}_i \) corresponds to a hypothesis concerning the state of the subject given the spatial relationship in \( \phi_i \) and the anchor, speaker, listener states etc. The natural ambiguity in spatially referring expressions and the difficulty in modelling such relationships in an autonomous way has been explored extensively in the language community; see e.g. [2]–[4], [18]. Our work differs from this existing work in that we explicitly want to model the likelihood function for a subject’s state given certain spatial relations using a rigorous mathematical framework that is well suited to the modelling problem at hand and is further tailored for a recursive fusion algorithm.

A. Discussion

In the next section we outline a rigorous mathematical framework for generally modelling the likelihood function for a subject’s state given certain spatial relations. We will also come back to the modelling problem later for some example spatial relations.

However, we note that the functional properties of the components involved, e.g. the anchors, in certain propositions has been attributed to human’s ability to naturally disambiguate certain spatially referring expressions [4]–[6], [18]. The proposition

\[
\phi_i = \text{the lightbulb is in the socket}
\]

is a typical example where the functional relationship between the lightbulb and the socket implicitly implies the geometrical relationship.

Certain expressions, like the one given in Example 2, cannot be easily disambiguated even by humans without additional information. For example, the proposition

\[
\phi_i = \text{the man is at the shop}
\]

may imply that the man is inside the shop or near the shop where inside and near may also be individually ambiguous. Nevertheless, such propositions are quite natural. Most likely, a listener attempting to find the target in Example 2 can quickly search a number of plausible hypotheses and eliminate them via an inherent recursive fusion algorithm, e.g. a recursive Bayes estimator, with vision essentially nullifying certain modes in the posterior. This is also the idea behind various multi-modal-based language interpretation systems; e.g. see [9], [10], [19]–[21]. On the other hand, additional propositions \( \phi_j \) concerning the location of the target may achieve the same outcome. This latter scenario is the one explored here and the main topic of this work.

III. GENERALIZED LIKELIHOOD FUNCTIONS FOR SPATIALLY REFERRING EXPRESSIONS

In this section we construct a likelihood function of the form

\[
g(\phi_i|\Sigma) \triangleq (1 - \varphi_i) + \sum_{k_i \in \mathcal{H}_i} w_{k_i} \gamma_{k_i}(\phi_i, \cdot)
\]

where \( \sum_{k_i} w_{k_i} = \varphi_i \) and explore the nature of this function in relation to traditional Bayesian estimation and information fusion. Again \( \varphi_i \in [0, 1] \) is the associated confidence, or truth value, of \( \phi_i \).

To this end, fix the underlying probability space \( (\mathcal{S}, \mathcal{B}(\mathcal{S}), P) \) where \( \mathcal{B}(\cdot) \) is a Borel \( \sigma \)-algebra. Let \( S^* \) denote the set of all closed subsets of \( \mathcal{S} \) equipped with the Mathéron, or hit-and-miss, topology [11]. We introduce the measurable space \( (S^*, \mathcal{B}(S^*)) \). A random closed subset \( \mathcal{X} \) of \( \mathcal{S} \) is a random element, generalizing the notion of a real-valued random variable, that is defined by a measurable map \( \mathcal{X} : \mathcal{S} \to S^* \). The push-forward probability measure of a random set \( \mathcal{X} \) is

\[
P_\mathcal{X}(A) = P(\{\cdot \in \mathcal{S} : \mathcal{X}(\cdot) \in A\}) = P(\mathcal{X}^{-1}(A))
\]

for \( A \in \mathcal{B}(S^*) \).

It is useful to think of \( \phi_i \) as a map \( \phi_i : \mathcal{S} \to \mathcal{R} \times \mathcal{I} \) taking the state of the subject in \( \mathcal{S} \) to the spatial relation \( \mathcal{R}(\phi_i) \) and anchor \( A(\phi_i) \) in the space \( \mathcal{R} \times \mathcal{I} \). The speaker, listener states etc are like parameters. Owing to the vagueness, ambiguity, imprecision etc in the spatial relationship referenced in \( \phi_i \) it is not typically true that \( \phi_i^{-1} \) should map to a singleton \( \Sigma \) in the state space \( \mathcal{S} \). Therefore, it is useful to think of the inverse proposition \( \phi_i^{-1} \) like a map \( \phi_i^{-1} : \mathcal{S}^* \to S^* \) taking the spatial relationship, anchor, speaker, listener states etc to one or more elements of \( \mathcal{S}^* \). For this reason we model \( \phi_i^{-1} \) as a realization of a random set \( \Phi_i^{-1} \) and we can define the generalized likelihood function by

\[
g(\phi_i|\Sigma) \triangleq (1 - \varphi_i) + \varphi_i P(\Sigma^{-1}(\Phi_i^{-1}))
\]

\[
= (1 - \varphi_i) + \varphi_i P(\Phi_i^{-1})(12)
\]

Note the likelihood function given \( \phi_i \) is modelled as a function over the state space \( \mathcal{S} \) even though \( \phi_i^{-1} \) is modelled as a realization of random set \( \Phi_i^{-1} \) in the space \( S^* \).
If one defined a measurable space \((\mathcal{X}, \mathcal{B}(\mathcal{X}))\) where \(\mathcal{X}\) is the set of all closed subsets of \(\mathcal{X}\) and then defines a random set \(\Phi_i : \mathcal{S} \rightarrow \mathcal{X}\) we find

\[
g(\phi_i | \Sigma) \triangleq (1 - \varphi_i) + \sum_{k_i \in H_i} w_{k_i} P(\{\Sigma\} \cap \Phi_{k_i} \neq \emptyset) \tag{13}
\]

is equivalent to the expressions in (12) in a more abstract language of spatial relations. It is easier to work with (12).

Suppose instead of a single random set \(\Phi_{k_i}^{-1}\) we use a finite number of random sets \(\Phi_{k_i}^{-1}, k_i \in H_i\) and define the likelihood function by

\[
g(\phi_i | \Sigma) \triangleq (1 - \varphi_i) + \sum_{k_i \in H_i} w_{k_i} P(\{\Sigma\} \cap \Phi_{k_i}^{-1} \neq \emptyset) \tag{14}
\]

with \(\sum_{k_i} w_{k_i} = \varphi_i\). Suppose \(P(\Phi_{k_i}^{-1} = \Phi_{k_i,ji}^{-1}) = 0\) for all but a finite number of \(\Phi_{k_i,ji}^{-1} \in \mathcal{S}^*\). Then

\[
P(\{\Sigma\} \cap \Phi_{k_i}^{-1} \neq \emptyset) = \sum_{j_i \in \varrho_{k_i}} v_{k_i,j_i} P(\Phi_{k_i}^{-1} = \Phi_{k_i,ji}^{-1}, \Sigma \in \Phi_{k_i,ji}^{-1}) \tag{15}
\]

where \(\sum_{j_i} v_{k_i,j_i} = 1\). Define a kind of (fuzzy) membership function \(m : \{S, \emptyset\} \rightarrow [0, 1]\) such that \(m(\emptyset) \equiv 0\). Let \(\mathcal{M}(\mathcal{S})\) denote the set of all functions \(m(\cdot)\) on \(\mathcal{S}\). Then

\[
P(\{\Sigma\} \cap \Phi_{k_i}^{-1} \neq \emptyset) = \sum_{j_i \in \varrho_{k_i}} v_{k_i,j_i} P(\Phi_{k_i}^{-1} = \Phi_{k_i,ji}^{-1}, \Sigma \in \Phi_{k_i,ji}^{-1}) = \sum_{j_i \in \varrho_{k_i}} v_{k_i,j_i} m_{k_i,j_i}(\|\Phi_{k_i,ji}^{-1}(\Sigma)\|) \tag{16}
\]

where \(\mathcal{J}_A(\cdot) : \mathcal{S} \rightarrow \{0, \cdot\}\) is a kind of indicator function that returns its argument \((\cdot)\) which is some subset of \(\mathcal{S}\) if this argument is a subset of \(A\) or else it returns \(0\). For example, \(\mathcal{J}_{\Phi_{k_i,ji}}^{-1}(\Sigma)\) returns \(\Sigma\) if \(\Sigma \in \Phi_{k_i,ji}^{-1}\) or else it returns \(\emptyset\). Then \(m_{k_i,j_i} \in \mathcal{M}(\mathcal{S})\) is defined over \(\{S, \emptyset\}\) and specifies a membership value for \(\Sigma \in \Phi_{k_i,ji}^{-1}\).

Finally

\[
g(\phi_i | \Sigma) \triangleq (1 - \varphi_i) + \sum_{k_i \in H_i} \sum_{j_i \in \varrho_{k_i}} v_{k_i,j_i} m_{k_i,j_i}(\|\Phi_{k_i,ji}^{-1}(\Sigma)\|) \triangleq (1 - \varphi_i) + \sum_{k_i \in H_i} w_{k_i} \gamma_{k_i}(\varphi_i, \cdot) \tag{17}
\]

with \(\sum_{k_i} w_{k_i} = \varphi_i\). We restrict ourselves to likelihood functions that can be modelled in such a form. The random set based nature of this likelihood function means that a large class of generalized likelihood functions can be modelled in such a form. In particular, the Bayesian fusion algorithm based on generalized likelihood functions of this form subsumes as a special case, Dempster-Shafer theory, fuzzy set theory, Bayesian fusion with likelihood mixture models etc.

IV. BAYES ESTIMATOR

Suppose there exists a set \(\{\phi_i\}_{i=0}^t\) of propositions concerning the value of the state \(\Sigma\). Then using Bayes formula

\[
p(\Sigma | \{\phi_i\}_{i=0}^t) = \frac{\int_S g(\phi_i | \Sigma) p(\Sigma | \{\phi_i\}_{i=0}^t - 1)} {\int_S p(\Sigma | \{\phi_i\}_{i=0}^t - 1)} \tag{18}
\]

where \(\{\phi_0\} \triangleq \emptyset\) and \(p(\Sigma | \{\phi_i\}_{i=0})\) is the defined prior probability for \(\Sigma\) on \(\mathcal{S}\). Note \(t \in \mathbb{N}\) may, or may not, index time, e.g. if the propositions concerning \(\Sigma\) come sequentially over a period of time.

A. A Recursive Particle Filter

A numerical solution based on the particle filter [16], [22]–[24] is proposed. The key idea of particle filters is to approximate the posterior \(p(\Sigma | \{\phi_i\}_{i=0}^t)\) by a set of random samples (particles) in a recursive manner; e.g. as new propositions become available or as the target evolves according to some known (but possibly uncertain) model. Thus, the posterior \(p(\Sigma | \{\phi_i\}_{i=0}^t)\) at time \(t - 1\) is approximated as:

\[
p(\Sigma | \{\phi_i\}_{i=0}^t - 1) \approx \sum_{i=0}^t c_{t-1}(i) \delta_{x_{t-1}(i)}(\Sigma) \tag{19}
\]

where \(x_{t-1} \in \mathcal{S}\) are the particles and \(c_{t-1}(i)\) are their associated weights. We suppose that \(\Sigma\) is stationary in \(\mathcal{S}\) over time \(t \in \mathbb{N}\) then each particle is updated using the measurement likelihood function input at \(t\).

This implementation assumes the speaker order, i.e. the order in which the \(\phi_i\) are applied, is irrelevant. This assumption is valid in this work since there is no temporal aspect of the individual statements.

B. Discussion

The problem considered here involves the fusion of multiple propositions concerning the state of a subject with the aim of increasing ones knowledge about the state value. Intuitively, one is seeking to reduce the number of modes in \(p(\Sigma | \{\phi_i\}_{i=0}^t)\) and increase the sharpness of one particular mode.

Although we focus on measurements that come as natural language propositions it is straightforward to include addition measurement types in such an algorithm. For example, humans inherently use vision to reduce the uncertainty/ambiguity produced by an ambiguous spoken proposition.

V. EXAMPLE LIKELIHOOD CONSTRUCTIONS

We outline certain parametrized likelihood functions for various common spatial relationships \(\mathcal{R}(\phi_i)\) in a target localization/positioning scenario given the anchor \(A(\phi_i)\). The state of the speaker \(s \in \mathbb{R}^2 \times S^1\) and listener \(\mathcal{P} \in \mathbb{R}^2 \times \Sigma^1\) are points on the plane \(\mathbb{R}^2\) with orientations. The state of the anchor is either a point with an orientation \(A(\phi_i) = a \in \mathbb{R}^2 \times S^1\) or a closed region \(A(\phi_i) = a \subset \mathbb{R}^2\) of the plane.

Denote the distance in \(\mathbb{R}^2\) between \(\mathbf{x}\) and \(y\) by \(d_{\mathbf{x},y} = \|\mathbf{x} - \mathbf{y}\|\). If \(\mathbf{x}\) and \(\mathbf{y}\) are in \(\mathbb{R}^2 \times S^1\) then \(d_{\mathbf{x},y}\) is the distance between the points neglecting the orientation. The distance between a point \(\mathbf{x} \in \mathbb{R}^2\), or \(\mathbf{x} \in \mathbb{R}^2 \times S^1\) when neglecting the
orientation, and a set $A \subseteq \mathbb{R}^2$ is $d_{x,A} = \inf\{d_{x,y} : y \in A\}$. Let $\mathbb{I}_A(\cdot)$ denote the standard indicator function.

We denote the ray defined by $x, y \in \mathbb{R}^2$ by $\ell_{x,y}$ and note that $x$ is the initial point of such a ray and $y$ is a point a finite distance $d_{x,y} < \infty$ away on $\ell_{x,y}$. Each ray is thus a line segment that is finite in one direction and infinite in the other. If $x$ and $y$ are in $\mathbb{R}^2 \times S^1$ then $\ell_{x,y}$ is the ray defined by neglecting the orientation. If $x$ is in $\mathbb{R}^2 \times S^1$ then $\theta_0$ is the ray starting at the $\mathbb{R}^2$ location of $x$ and heading in the direction $\theta$ taken positive-counter-clockwise from the orientation of $x$.

A. Relationship: $\mathcal{R} = \{\text{near}\}$

We consider the spatial relationship $\mathcal{R} = \{\text{near}\}$ with $a \in \mathbb{R}^2 \times S^1$. The first likelihood considered is

$$g(\phi|\Sigma) \triangleq (1 - \varphi_i) + \varphi_i \gamma(\Sigma - a, \Omega)$$

where $\gamma(x - \mu, \Sigma)$ is a Gaussian density (7) and $\Omega$ is a tuning parameter. It would be typical to tune $\Omega$ based on something like the distance $d_{a,n}$ between $a$ and $s$. For example, the closer the speaker is to the anchor, the closer the target would be to the anchor and thus the smaller $\Omega$ should be.

Another likelihood for $\mathcal{R} = \{\text{near}\}$ with $a \in \mathbb{R}^2 \times S^1$ can be defined by first supposing $H_i = \{1, \ldots, e_i\}$ with $e_i < \infty$. Then denote the sets

$$\Phi^{-1}_i \subset \Phi^{-2}_i \subset \ldots \subset \Phi^{-e_i}_i \subset \mathbb{R}^2$$

where $\Phi^{-1}_i$ is a disk in $\mathbb{R}^2$ centered at $a \in \mathbb{R}^2 \times S^1$ with radius $d_k$ and $d_1 < d_2 < \ldots < d_{e_i}$. These $d_k$ are tuning parameters. Then

$$g(\phi|\Sigma) \triangleq (1 - \varphi_i) + \sum_{k_i \in H_i} w_{k_i} \mathbb{I}_{\Phi^{-k_i}_i}(\Sigma)$$

where $\sum_{k_i \in H_i} w_{k_i} = \varphi_i$. Since $\Phi^{-1}_i$ are realizations of a random set $\Phi_k$, the tuning parameters $w_{k_i}$, if $\varphi_i = 1$, can be interpreted via $P(\Phi^{-k_i}_i = \Phi^{-1}_i) = w_{k_i}$. The radii $d_1 < d_2 < \ldots < d_{e_i}$ should typically be a function of $d_{a,n}$. For example, $d_1 < d_{a,n}$ makes sense for all $k_i$ except perhaps $k_i \leq e_i$ with $w_{e_i}$ sufficiently small. One approach is to set

$$d_{k_i} = k_i d_{a,n}/(e_i - 1) \text{ and } w_{k_i} = (\varphi_i - e)/(e_i - 1)$$

for $k_i \leq e_i - 1$. Then $d_{e_i} = \infty$ and $w_{e_i} = e$ for some small $e$.

Finally, we consider a likelihood for $\mathcal{R} = \{\text{near}\}$ when $a \in \mathbb{R}^2$ is some closed region of the plane. Suppose $H_i = \{1, \ldots, e_i\}$ with $e_i < \infty$. Then denote the closed sets

$$a \subset \Phi^{-1}_i \subset \Phi^{-2}_i \subset \ldots \subset \Phi^{-e_i}_i \subset \mathbb{R}^2$$

in $\mathbb{R}^2$ and define the closed sets $\Phi^{-1}_i = \Phi^{-1}_i/a$ where the notation $\partial a$ denotes the boundary of $A$. Then

$$g(\phi|\Sigma) \triangleq (1 - \varphi_i) + \sum_{k_i \in H_i} w_{k_i} \mathbb{I}_{\Phi^{-k_i}_i}(\Sigma)$$

where $\sum_{k_i \in H_i} w_{k_i} = \varphi_i$. The boundary $\partial a$ of a closed region $a \subset \mathbb{R}^2$ on the plane can be described by a closed curve and more specifically as a continuous mapping of the circle $S^1$. One intuitive way to define $\Phi^{-1}_i$ is by blowing up the closed curve defining $\partial a$ continuously such that $\Phi^{-1}_i$ is the same shape as $a$. The amount of blow up should be proportional to $d_{a,n} = \inf\{d_{a,x} : x \in a\}$ as before.

B. Relationship: $\mathcal{R} = \{\text{inside}\}$, $\mathcal{R} = \{\text{outside}\}$

We consider first the spatial relationship $\mathcal{R} = \{\text{inside}\}$ with $a \in \mathbb{R}^2$ being some closed region of the plane. Then

$$g(\phi_i|\Sigma) \triangleq (1 - \varphi_i) + \varphi_i \mathbb{I}_{\Phi^{-1}_i}(\Sigma)$$

and $\Phi^{-1} = a$. One could define a similar likelihood function for the spatial relationship $\mathcal{R} = \{\text{in}\}$ but care must be taken to ensure that this relationship $\mathcal{R} = \{\text{in}\}$ is used to convey physical containment and not a functional relationship, e.g. as in $\phi_i =$ the light bulb is in the socket. The relationship $\mathcal{R} = \{\text{inside}\}$ is perhaps stronger, in the sense of conveying physical containment, than $\mathcal{R} = \{\text{in}\}$. If we restrict ourselves to the target localization/positioning scenario than further assumptions can probably be made.

We consider now the spatial relationship $\mathcal{R} = \{\text{outside}\}$ with $a \in \mathbb{R}^2$ being some closed region of the plane. Then

$$g(\phi_i|\Sigma) \triangleq (1 - \varphi_i) + \varphi_i \mathbb{I}_{\Phi^{-1}_i}(\Sigma)$$

and $\Phi^{-1} = a$. One could define a similar likelihood function for the spatial relationship $\mathcal{R} = \{\text{outside}\}$ but care must be taken to ensure that this relationship $\mathcal{R} = \{\text{outside}\}$ is used to convey physical containment and not a functional relationship, e.g. as in $\phi_i =$ the light bulb is in the socket. The relationship $\mathcal{R} = \{\text{outside}\}$ is perhaps stronger, in the sense of conveying physical containment, than $\mathcal{R} = \{\text{outside}\}$. If we restrict ourselves to the target localization/positioning scenario than further assumptions can probably be made.

C. Relationship: $\mathcal{R} = \{\text{in front}\}$, $\mathcal{R} = \{\text{behind}\}$

We consider first the similar spatial relationship $\mathcal{R} = \{\text{in front}\}$ with $a \in \mathbb{R}^2 \times S^1$. The orientation of $a$ defines the front of the anchor in a natural way. The likelihood function is

$$g(\phi_i|\Sigma) \triangleq w_{1} \gamma_1(\Sigma - q_1, \Omega_1) + w_{2} \gamma_2(\Sigma - q_2, \Omega_2)$$

$$+ w_{3} \gamma_3(\Sigma - q_3, \Omega_3) + (1 - \varphi_i)$$

where $w_1 + w_2 + w_3 = \varphi_i$ and $\gamma(x - \mu, \Sigma)$ is a Gaussian density (7) and the means $q_{k_i}$ and variances $\Omega_{k_i}$ are tuning parameters. Each component $k_i \in H_i$ is motivated by the notion that $\mathcal{R} = \{\text{in front}\}$ may imply the target is in front of the anchor with respect to the speaker position, the listener position or the anchor orientation itself respectively.

The location of the three means $q_{k_i}$ would lie on $\ell_{a,n}$, $\ell_{a,p}$ and $\ell_{a,p}^0$ respectively for $k_i \in \{1, 2, 3\}$. The exact position of the mean values on $\ell_{a,n}$ and $\ell_{a,p}^0$ are dependent, in a natural way, on $d_{a,n}$ as are the relevant variances. The position of the mean value on $\ell_{a,p}$ is tuned based also on the distance $d_{p,a}$ as is the relevant variance.

Another likelihood for $\mathcal{R} = \{\text{in front}\}$ with $a \in \mathbb{R}^2 \times S^1$ can be defined by first setting $H_i = \{1, 2, 3\}$ and defining three disks $\Phi^{-1}_{k_i}$ in $\mathbb{R}^2$ centered at $a \in \mathbb{R}^2 \times S^1$ with radii $d_{k_i}$. For each $\ell_{a,n}$, $\ell_{a,p}$ and $\ell_{a,p}^0$ define two additional rays by rotating the relevant $\ell_{a,n}$, $\ell_{a,p}$ or $\ell_{a,p}^0$ positive-counter-clockwise by $\alpha_{k_i}$ and negative clockwise by $\alpha_{k_i}^{-1}$ about the anchor. Denote the subsequently defined conic sets subtended at the anchor by the angle $2\alpha_{k_i}$ by $\Phi^{-1}_{k_i}$. Define $\Phi^{-1}_{k_i} = \Phi^{-1}_{k_i} \cap \Phi^{-1}_{k_i}$ such that $\Phi^{-1}_{k_i}$ is a wedge-like set. Then

$$g(\phi_i|\Sigma) \triangleq (1 - \varphi_i) + \sum_{k_i \in H_i} w_{k_i} \mathbb{I}_{\Phi^{-k_i}_i}(\Sigma)$$

where $\sum_{k_i \in H_i} w_{k_i} = \varphi_i$. The radii $d_{k_i}$ and angles $2\alpha_{k_i}$ are tuning parameters that define the dimensions of the wedge-like sets $\Phi^{-1}_{k_i}$.
The spatial relation $\mathcal{R} = \{\text{behind}\}$ is similar in principle to $\mathcal{R} = \{\text{in front}\}$ and the relevant likelihood function for $\mathcal{R} = \{\text{behind}\}$ follows in an obvious way from the likelihood for $\mathcal{R} = \{\text{in front}\}$.

D. Relationship: $\mathcal{R} = \{\text{at}\}$

We consider the spatial relationship $\mathcal{R} = \{\text{at}\}$ with $a \in \mathbb{R}^2$ being some closed region of the plane. Suppose $\mathcal{H}_i = \{2, \ldots, e_i\}$ with $e_i < \infty$. Then denote the closed sets

$$a \subset \Phi_2^{-1} \subset \Phi_3^{-1} \subset \ldots \subset \Phi_{e_i}^{-1} \subseteq \mathbb{R}^2$$

in $\mathbb{R}^2$ and define the closed sets $\Phi_1^{-1}_k = \Phi_1^{-1}/\{a/\partial a\}$. Then

$$g(\phi_i|\Sigma) \triangleq (1 - \varphi_i) + w_1 \mathbb{I}_{\Phi_1^{-1}}(\Sigma) + \sum_{k_i \in \mathcal{H}_i} w_{k_i} \mathbb{I}_{\Phi_1^{-1}_i}(\Sigma)$$

with $w_1 + \sum_{k_i \in \mathcal{H}_i} w_{k_i} = \varphi_i$ and $\Phi_1^{-1}_i = a$. Typically, one would also assume $w_1 = \sum_{k_i \in \mathcal{H}_i} w_{k_i}$. Thus, for the spatial relation $\mathcal{R} = \{\text{at}\}$ the likelihood $g(\phi_i|\Sigma)$ is a combination of the likelihood functions for $\mathcal{R} = \{\text{inside}\}$ and $\mathcal{R} = \{\text{near}\}$. The modelling of $\Phi_1^{-1}_k$, $k_i \in \mathcal{H}_i$ should follow the same procedure as for the spatial relation $\mathcal{R} = \{\text{near}\}$ and would depend intuitively on $d_{a,a} = \inf\{d_{a,x} : x \in a\}$.

E. Discussion

There are spatial relationships which in many cases are homomorphic to the ones considered in this section; e.g. $\{\text{close to}\} \sim \{\text{near}\}$ etc. Some spatial relations are homomorphic to combinations of some relations considered in this section; e.g. $\mathcal{R} = \{\text{next to}\}$ is related to $\{\text{in front}\}$ and $\{\text{behind}\}$ where the front etc. of the anchor is ambiguous.

F. Disclaimer

We maintain that it is generally better to model the likelihood functions, given particular spatial relationships, robustly and then rely on the fusion of multiple propositions to reduce the uncertainty in the subjects state. The general nature of the likelihood $g(\phi_i|\Sigma) \triangleq (1 - \varphi_i) + \varphi_i P(\{\Sigma\} \cap \Phi_1^{-1} \neq \emptyset)$ given in (12) is based on the principle that a state in $\mathcal{S}$ should be consistent with the measurement information defined by the random set model $\Phi_i^{-1}$ so long as this state does not flatly contradict it. This is, by its own accord, a robust notion of model matching [12]. One then relies on the fusion of multiple statements (with differing parameters and spatial relationships) to reduce the uncertainty of the subject state.

In this section we proposed some example likelihood functions for commonly spoken spatial relations and described how certain tuning parameters may be determined in a target localization/positioning scenario. This exposition is by no means exhaustive nor are the functions defined for the individual examples the only possible choices.

We believe the generalized likelihood framework outlined in this paper is sufficient to model very complex likelihood functions for propositions in the form (1). This framework has an appealing intuitive aspect in the context of random set theory and is designed with robustness in mind.

VI. AN ILLUSTRATIVE EXAMPLE

We consider a simple illustrative example of target localization/positioning with spatially referring propositions. We use an implementation of a particle filter; see [16], [23], [24] for various particle filter implementations. Such an example is indicative of a realistic scenario in which information fusion, as detailed in this work, would be advantageous in practice. The scenario is depicted in Figure 2.

There are five speakers in the field of interest. The location of Speakers 1 and 2 in this example is unimportant, while Speaker 3 is located at (10,50), Speaker 4 is located at (80,10) and Speaker 5 is located at (50,10). There is a single listener whose location in this example is unimportant. Speaker 1 states firstly that:

$$\phi_1 = \text{the target is in the field}$$

with $\varphi_1 = 1$.

If we interpret this statement to mean the target is within the field proper, i.e. not in the building, garage or tower, then the initial posterior $p(\Sigma|\{\phi_1\}_{i=0})$ can be approximate using particles as in Figure 3.

The particles in Figure 3 are spread uniformly across the field. Speaker 2 then states:

$$\phi_2 = \text{I am pretty sure the target is near}$$

the garage or near the pool

and we interpret “I am pretty sure” to mean $\varphi_2 = 0.7$. The statement following “I am pretty sure” is in normal form. The updated posterior $p(\Sigma|\{\phi_2\}_{i=0})$ appears as in Figure 4.

The particles in Figure 4 are now more concentrated near the pool and the garage as expected. Speaker 3 then states:

$$\phi_3 = \text{I do not see the target}$$

with $\varphi_3 = 1$. This statement can be transformed into normal form via a homomorphism. For example, the statement

$$\phi_3 = \text{the target is outside the visibility polygon of speaker 3}$$
is homomorphic to the original statement. A visibility polygon is a well-defined geometric structure which in this case is a star-shaped polygon and can be found in linear time. The updated posterior $p(\Sigma|\{\phi_i\}_{i=0}^3)$ appears as in Figure 5.

The particles in Figure 5 are now evacuated from the visibility polygon of speaker 3 as expected. Speaker 4 then states:

$$\phi_4 \quad \text{the target is in front of the tower} \quad (35)$$

with $\varphi_4 = 1$. The updated posterior $p(\Sigma|\{\phi_i\}_{i=0}^4)$ appears as in Figure 6.

We note there are no more particles concentrated near the pool in Figure 6 since the previous statement essentially negates the hypothesis that the target may be near the pool. Finally, Speaker 5 states:

$$\phi_5 \quad \text{the target is at 1 o’clock} \quad (36)$$

with $\varphi_5 = 1$. Again, this statement is homomorphic to one in normal form. We define a two-dimensional cone-based uniform distribution with a $\pm 5^\circ$ spread centered at 1 o’clock (i.e. $30^\circ$ clockwise from north) with an apex at Speaker 5. The updated posterior $p(\Sigma|\{\phi_i\}_{i=0}^5)$ appears as in Figure 7.

In the final posterior $p(\Sigma|\{\phi_i\}_{i=0}^5)$ depicted in Figure 7 we note the relative accuracy we have achieved in locating the target given nothing but rather vague individual statements about
its possible location. Through the fusion of these statements we have shown how we can refine our knowledge concerning the property of the subject in question, e.g. in this case the position of the target. While this example was conceived to illustrate the principle underlying this work it is by no means unrealistic and the construction of the likelihood functions and the posterior probabilities was entirely realistic.

VII. CONCLUDING REMARKS

An automatic information fusion scheme was introduced for state estimation where the inputs (or measurements), that are used to reduce the uncertainty in the state of a subject, are in the form of natural language propositions. A mathematically rigorous method to generate likelihood functions using natural language propositions was developed using the framework of random-set-based probability. We argued that one should model such likelihood functions robustly and account for the natural ambiguity and uncertainty in the propositions. One then relies on the fusion of multiple statements (with differing parameters and spatial relationships) to reduce the uncertainty of the subject state. A recursive Bayesian algorithm was outlined to this end and an illustrative example was provided.

REFERENCES


VIII. APPENDIX

Combinatory categorial grammar (CCG) is an efficiently parseable, yet linguistically expressive grammar formalism and it is used as the basis for a language parser in this work to justify the normal form (1). It has a transparent interface between surface syntax and underlying semantic representation, including predicate-argument structure, quantification and information structure [25], [26].

For example, suppose we are given two statements:

\[ \phi = \text{the ball is near the door} \]
\[ \phi = \text{near the door is the ball} \]

Then a parsing of either statement using an open source CCG implementation might give something like:

```plaintext
@c1:event (context ^
  <Tense> present ^
  <Modifier>(n1:location ^ near ^
    <Anchor>(d1:thing ^ door ^
      <Delimitation> unique ^
      <Num> singular ^
      <Quantification> specific)) ^
    <Subject>(b1:thing ^ ball ^
      <Delimitation> unique ^
      <Num> singular ^
      <Quantification> specific))
```

The semantic parsing of each statement describes an event, or more specifically a context, in which something (the subject, i.e. ball) is in a location that is near (the spatial relationship) an anchor (i.e. the door).

The advantage of working with semantics rather than with syntactic structures is that semantics are much more invariant. That is, you can express the same meaning in many different ways. This type of semantic parsing provides the basis for the normal proposition form (1) used as input to the fusion algorithm described in this work.