Abstract – This paper presents a novel approach to data fusion knowledge representation using conceptual spaces. Conceptual spaces represent knowledge geometrically in multiple domains, each domain consisting of multiple dimensions with an associated distance metric and corresponding similarity measure. Complex concepts such as those required for Level 2/3 fusion are described by multiple property regions within these domains, along with the property correlations and salience weights. These concepts are mapped into points in the unit hypercube that capture all of their essential elements. Observations are also mapped into points in the same unit hypercube. The relative similarity of observations to concepts can then be calculated using the fuzzy subsethood measure.

Keywords: Conceptual spaces, knowledge representation, fuzzy systems, inference mechanisms.

1 Introduction

Level 2 fusion is concerned with situation understanding, while Level 3 fusion is concerned with threat assessment. Both of these fusion levels rely upon knowledge representations and inference mechanisms that have relatively little in common with the algorithmic approaches typically employed in Level 1 fusion (object refinement).

One of the principal impediments to progress in Levels 2 and 3 fusion is the availability of an appropriate knowledge representation. The two dominant modalities are the symbolic and connectionist representations. The symbolic approach describes basic knowledge elements using symbols, employs propositional statements (i.e., rules) to capture the relationships between symbols, and thus reduces fusion to computation in the form of symbol manipulation. The connectionist (i.e., neural network) approach typically deals with much lower level representations of knowledge (often in the form of raw data or low level features with little if any semantic significance), and relies upon the learning of parallel nonlinear associations between these data elements for performing fusion tasks.

While both of these approaches have their strengths, there is a considerable gulf between them that leaves unaddressed many important aspects of cognition required to perform Level 2/3 fusion tasks. In this paper, we propose a novel approach to Level 2/3 knowledge representation using the conceptual spaces (CS) approach of Gardenförs [1], with further extensions developed by this author [2]. We also describe an inference mechanism based upon a fundamental measure from fuzzy set theory that operates upon this representation, and illustrate its use with a Level 2 fusion example.

2 Conceptual Spaces

The original motivation for conceptual spaces was to provide an eclectic approach to knowledge representation that exploits the geometric structures embodied in many of the attributes of the real world, while preserving a semantically meaningful description of these attributes. Thus conceptual spaces lie midway between the symbolic and connectionist approaches.

2.1 Dimensions

From Gardenförs [1], the starting point for a conceptual space is a set of dimensions capable of describing the quality attributes of the information to be represented. These dimensions can be either psychophysical (which measure human phenomenal responses and are thus semantically meaningful within the constructs of natural language) or scientific (which measure the values associated with sensors, actuators, etc.) For a given application, there is generally no unique assignment of dimensions. Instead, those who are most familiar with the application (often referred to as “subject matter experts” or SMEs) will specify the appropriate dimensions that capture its essential qualities. This represents the bare minimum of “knowledge engineering” required to model a problem.

Dimensions generally possess geometric and/or topological structures that enable us to measure distances between two values. In the case of continuous linear variables such as time or spatial coordinate axes, the distance is typically a function of the difference between the two values. However, many variables have more complex distance functions, e.g., where the underlying quality has a circular structure, in which case the distance might be specified as the length of an arc generated by convex combinations in both the radius and angle dimensions, using the same parameter to trace the arc. We can also envision discrete structures such as graphs.
where the distance between two nodes is measured by the length of the shortest path connecting them.

### 2.1 Domains

Dimensions are organized into multiple domains, which need not be Cartesian coordinate systems. The defining feature of a domain is that its dimensions are integral, in the sense that a logical distance measure incorporating all of the domain dimensions can be assigned within the domain, whereas no such measure can be assigned across domains. In the conceptual spaces familiar to humans, some domains are innate and apparently hard-wired in our sensory apparatus, e.g., color, aural pitch and ordinary three-dimensional space. Other domains are learned or culturally dependent, while still others are induced by science. The ability to add domains to a conceptual space further to define information regarding objects of interest endows it with virtually unlimited capability for knowledge representation.

Objects in a conceptual space are represented by points or sets of points in each domain that characterize their dimensional values. This constitutes the fundamental geometric character of conceptual spaces for knowledge representation. Within each domain, we can measure the distance between two objects as the distance between their corresponding points or sets.

Inherent to the notion of distance is the notion of similarity. Two values are considered to be similar to the extent that the distance between them is small, and vice versa. The distance \( d(x,y) \) between two points \( x \) and \( y \) may be measured directly using, for example, the Euclidean distance metric, while the distance between sets is typically measured using the Hausdorff metric. Similarity is generally described as a fuzzy function of distance \( s(x,y) = f[d(x,y)] \) taking values in the interval \([0,1]\), which expresses the fuzzy degree to which the two points \( x \) and \( y \) are similar. If \( d(x,y) \leq \delta \) for some \( \delta \geq 0 \), the similarity \( s(x,y) \) is considered to be unity, while if \( d(x,y) \geq \gamma \) for some \( \gamma > 0 \), \( s(x,y) \) is considered to be zero. For \( \delta < d(x,y) < \gamma \), \( s(x,y) \) takes on values between zero and unity, declining monotonically with increasing \( d(x,y) \). By working with fuzzy values for similarity, we can take advantage of concepts from fuzzy set theory to provide a linguistic interpretation of results, while preserving the essential geometric aspects of our knowledge representation.

### 2.2 Properties

A property is a convex region in some domain. The notion of convexity for property regions arises from the logical assumption that if two objects \( v_1 \) and \( v_2 \) possess some property, then objects located between \( v_1 \) and \( v_2 \) should likewise possess that property. The “betweenness” in some domains may not correspond to “lying on a line between”, e.g., in a domain possessing a circular dimension, we might use a joint convex combination of radius and angle coordinates to define an arc whose locus of points is “between” \( v_1 \) and \( v_2 \). In natural languages, properties often correspond to adjective-like descriptions (e.g., “red”, “tall”, or “round”) in a particular domain. Properties can also capture more complex descriptions of objects, including shapes, actions and functional characteristics, as described in [1, pp. 94-99]. They can also be defined in probabilistic or fuzzy terms, which provide a distinct advantage over representational schemes that require strict set membership.

The notion of properties is related to prototype theory, where certain members of a category of objects are considered to be most representative (i.e., prototypes) of the category as a whole, and to clustering theory, where the centroid of a cluster is taken to be representative of the cluster members.

### 2.3 Concepts

A concept is a combination of properties, typically across multiple domains, along with the salience weights associated with each property and the correlations (used in the sense of conditional similarities, as opposed to statistical correlations, which can be positive or negative) between properties. The choice of properties is predicated upon the descriptive features of the application. The salience weights may be dependent upon the context.

The distinction between concepts and properties is often obliterated in other representational schemes, which lump properties into predicates, classes or low-level features. Concepts typically describe noun- or verb-like objects (when time is included as a dimension in the latter case), whereas properties define the attributes of these objects in a geometric context.

On the surface, concepts bear some resemblance to the popular representational scheme of frames with slots for different features as originally proposed by Minsky [3]. However, as described below, the geometric structure of concepts within the conceptual space framework enables us to calculate the similarity of observations to concepts in a fuzzy sense, which is exceedingly difficult to implement in a symbolic system. We also note that the geometric structure of conceptual spaces provides an alternative way of organizing knowledge to the more hierarchical, taxonomic approach used in ontologies, and one that lends itself to a more analytical approach to representation and inference.

### 3 Concept Representations

The preceding discussion summarizes Gardenfors’ description of conceptual spaces. In the following sections, we present extensions to Gardenfors’ theory, which enable computable inferences upon multidimensional observations. The essential features of this approach are to construct a graphical representation of both concepts and observations, and then to transform these graphs into vector points in a higher dimensional unit hypercube. Portions of this material draw upon the author’s previous work in [2, 4] but we provide additional
insights into the interpretation of concepts, and the applications to Level 2/3 fusion problems are novel.

### 3.1 Graphical Representations

We can capture the elements of a concept in an attributed graph $C$, where each node in the graph corresponds to a property $P_i$ with its corresponding salience weight $w_j$ as an attribute. The weighted edges between nodes are determined by the domain structure of the concept. We assign directional edges between all pairs of nodes corresponding to properties having non-zero conditional similarities. Associated with each such edge is a directional edge strength $C_{ij}$ which, in the case of crisp property memberships, equals the fraction of an ensemble of object exemplars for the concept having property $P_i$ that also have property $P_j$. In this case, $C_{ij}=1$ if and only if all objects having property $P_i$ in one domain also have property $P_j$ in another, but in general $C_{ij} \neq C_{ji}$, $i \neq j$. We set $C_{ij} = 0$ if the pair $(i,j)$ corresponds to disjoint properties within the same domain; otherwise, we treat these values the same as for cross-domain properties. Note that concept connection matrices are square, but generally non-symmetric, and that cases where $C_{ij} = 0$ but $C_{ji} \neq 0$ happen only in theory (e.g., all unicorns have one horn, but no one-horned animals are unicorns).

When observations are assumed to arise from random processes, $C_{ij}$ is given by the expected value of the following similarity ratio:

$$C_{ij} = \mathbb{E}_{x_{ij}} \left[ \frac{s_{ij}(x_{ij})}{s_j(x_j)} \right],$$  \hspace{1cm} (1)$$

where $s_{ij}(x_{ij})$ is the joint similarity function of the augmented vector $x_{ij}=[x_i, x_j]^T$ to properties $P_i$ and $P_j$, $s_j(x_j)$ is the similarity function of $x_j$ to property $P_j$, and the expectation is taken over the joint distribution of $x_{ij}=[x_i, x_j]^T$. For example, if these similarity functions take the form of the exponential kernels of multivariate Gaussian densities with respective means and covariance matrices $(\mu_{ij}, A_{ij})$ and $(\mu_j, A_j)$, then

$$C_{ij} = \mathbb{E}_{x_{ij}} \left[ \frac{\exp \left[ -\frac{1}{2} ((x_i - \mu_{ij})^T A_{ij} (x_i - \mu_{ij})) \right]}{\exp \left[ -\frac{1}{2} ((x_j - \mu_j)^T A_j (x_j - \mu_j)) \right]} \right]$$

$$= \mathbb{E}_{x_{ij}} \left[ \frac{\exp \left[ -\frac{1}{2} ((x_i - \mu_j)^T A_j (x_i - \mu_j)) + (x_i - \mu_j)^T A_{ij} (x_j - \mu_j) \right]}{\exp \left[ -\frac{1}{2} (x_j - \mu_j)^T A_j (x_j - \mu_j) \right]} \right]$$

Note that the Gaussian similarity function does not suggest a Gaussian distribution of observations—it is simply the similarity function of the property. Even if the observation densities are multivariate Gaussian, they may in general possess different means and covariances from those used in these similarity functions. However, in the case where the latter are identical and where $x_i$ and $x_j$ are statistically uncorrelated (i.e., $A_{ij} = 0$), we would have

$$C_{ij} = \mathbb{E}_{x_i} \left[ \exp \left[ -\frac{1}{2} ((x_i - \mu_j)^T A_j (x_i - \mu_j)) \right] \right]$$

$$= (2\pi)^{-\frac{k}{2}} |A_j|^\frac{1}{2} \int dx_i \exp \left[ -\frac{1}{2} (x_i - \mu_j)^T A_j (x_i - \mu_j) \right]$$

$$= \frac{\sqrt{2}}{2}$$

where $k$ is the number of dimensions of domain $\Delta_j$. In other words, when the ensemble data are independently normally distributed between the two domains and the similarity functions have the same form of normal density kernels as the distribution functions of the data, then

$$C_{ij} = C_{ji} = \frac{\sqrt{2}}{2} = 0.707.$$ This is intuitive, since it simply says that the data distributions correlate well to each property. On the other hand, if $p_{x_i}(x_{ij})$ has little mass overlapping $s_j(x_j)$ in domain $\Delta_j$, then no matter how well the data distribution matches the similarity function $s_j(x_j)$ in domain $\Delta_j$, $C_{ij}$ will be small.

Our approach also accommodates fuzzy property similarity values and the resulting correlations. In this case, $C_{ij}$ represents the degree of similarity of objects in an ensemble to both properties $P_i$ and $P_j$ (i.e., the minimum of the respective similarity values) relative to the degree to which they are similar to property $P_j$.

Correlation matrices can be learned straightforwardly, given an exemplar data set, or in its absence they can be specified a priori. Suppose we are given a set of parameterized individual property similarity functions $p_j(x_j; \theta_j)$ and joint property similarity functions $s_{jk}(x_j, x_k; \theta_j, \theta_k)$ for $j \in \{1...N\}$, and a training set of observations $z_{ij}, i = 1...M$ for each of the properties representing concept $C$. For each training observation, we calculate its vector of individual similarities $q_i$ in the properties in concept $C$:

$$q_i = \left[ q_{i1}, q_{i2}, \ldots, q_{iN} \right]^T$$

$$= \left[ s_{1}(z_{i1}; \theta_1) \ s_{2}(z_{i2}; \theta_2) \ \ldots \ s_{N}(z_{iN}; \theta_N) \right]^T$$  \hspace{1cm} (2)$$

Thus the $j^{th}$ element of this vector represents the similarity of the $i^{th}$ training observation to the $j^{th}$ property in $C$. We further calculate the matrices of
pairwise joint similarities $R_i$ of the $i^{th}$ training
observation in the $j^{th}$ and $k^{th}$ properties in concept $C$:

\[ R_i(j,k) = s_{jk}(z_{ij}, z_{ik}; \theta_j, \theta_k) \]  

(3)

We then calculate the concept matrix for concept $C$ as

\[ C_{jk} = \frac{1}{M} \sum_i R_i(j,k) q_{ik}, \]  

(4)

where the summand is zero for zero values of $q_{ik}$. This
expression can be viewed as the average fuzzy subsethood
of the training set’s similarity to the $j^{th}$ property with
respect to its similarity to the $k^{th}$ property.

In constructing a decision system that must decide
between different concepts based upon observations, it is
often possible to specify an objective function for
measuring the efficacy of the similarity measures and
concept matrices resulting from a particular choice of
parameters $\theta$ over the training data set. One can then use
an evolutionary optimization algorithm over the parameters $\theta$ in the similarity measures to determine an
optimal set of parameters for maximizing this objective
function.

Figure 1 illustrates a notional concept matrix for a
conceptual space consisting of four domains. The
concept possesses two properties in each of the first two
domains, and one property in each of the last two
domains.

<table>
<thead>
<tr>
<th>Domain</th>
<th>Property 1</th>
<th>Property 2</th>
<th>Property 3</th>
<th>Property 4</th>
<th>Property 5</th>
<th>Property 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1  0  0.9  0  1  0.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0  1  0.7  0.4  1  0.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1  0.7  0.3  1  0.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0  0.4  0  1  1  1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.6  0.4  0.5  0.3  1  0.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>1  0.3  0.1  0.2  0.2  1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 1 An example concept matrix.

For example, the first domain might represent
kinematics, and each property would correspond to a
choice of routes that a group of contacts might take to an
objective. The second domain might represent group
compositions, with each property corresponding to a
different composition. The third and fourth domains
might represent group formation and region of influence,
respectively, with one property each. This example
illustrates how knowledge that is pertinent to situation
understanding (Level 2 fusion) and threat assessment
(Level 3 fusion) can be represented as a concept in a
conceptual space framework. One can also easily see
how the representation of the above concept can be
extended to include additional domains and properties,
and how other concepts can be formulated in similar
fashion.

3.2 Concepts as points

We employ Rickard and Yager’s hypercube graph
representation [4] to transform the connection matrix $C$
to a point in the $N^2$-dimensional unit hypercube, where $N$ is the matrix dimension. This is accomplished
by “unwinding” the matrix rows into a concept vector $c$, where each element of the vector corresponds to the
relation strength between a pair of properties, as
illustrated in Figure 2. Thus the graph represented by
the matrix $C$ now becomes a point $c$ in the $N^2$-
dimensional hypercube, analogous to a fuzzy set with
corresponding membership coordinates in each of the
latter dimensions. This hypercube description allows us
to invoke fuzzy set theoretic concepts for graph
representation and characterization, as described in [4].

\[
C = \begin{bmatrix}
1 & C_{12} & \cdots & C_{1N} \\
C_{21} & 1 & \cdots & C_{2N} \\
\vdots & \vdots & \ddots & \vdots \\
C_{N1} & C_{N2} & \cdots & 1 \\
\end{bmatrix}
\]

Figure 2 Representation of a connection matrix as a
point in the unit hypercube.

The salience weight $\omega_k$ of each property pair can be
taken, e.g., as the product of the individual property
salience weights: $\omega_k = w_i w_j$, $k \leftrightarrow (i, j)$, where the
vector index $k$ corresponds to the property pair $(i, j)$ as
above (other weighting combinations for the salience
weights such as the minimum, or any number of different
averaging operators, are also possible.)

Thus we have translated Gardenfors’ definition of a
concept into a point in the unit hypercube, which further
embellishes the geometric structure of conceptual spaces,
and more importantly, enables us to treat concepts as
multi-dimensional fuzzy sets in the space of property
pairs. Along with the point $c$, we have a vector $\omega$ of
salience weights associated with the property pairs.

4 Concept/Observation Similarity

In Level 2/3 fusion, we wish to compare observations
about the real world to a library of concepts stored in our
cognitive systems in order to perform inferences upon
these observations. A concept is generally more complex
than an observation, since concepts encode the
relationships among all possible combinations of
properties between domains, whereas a specific
observation corresponds to a single point or set of points
in each domain. Thus a concept describes the ensemble
properties of all objects that are representative of the
concept, while an observation is a single realization
corresponding to a particular object. In order to calculate
the similarity of an observation to a concept, we must cast
the former in the same geometric space as the latter, and
then apply an appropriate measure for this purpose.
4.1 Object representation

To relate an observation \( x \) (i.e., a set of vector points in each domain of the conceptual space) to a concept, we first calculate its similarity \( s_j = s(x, P_j^i) \) \((0 \leq s_j \leq 1)\) to each property \( P_i \) involved in the concept, using the similarity measure derived from the distance metric associated with each domain. Typically this involves calculating the domain-specific distance between \( x \) and the centroid (i.e., prototype point) of the property region \( P_i \) and then mapping this distance value into the normalized similarity measure.

In a given domain \( D_n \) of a conceptual space containing multiple properties \( P_j, j \in D_n \), a particular observation \( x \) will have various similarities \( s_j, j \in D_n \) to the properties \( P_i \) within that domain. Between two different domains, the minimum function \( \min(s_j, s_k) \) describes the fuzzy degree to which observation \( x \) possesses a pair of properties.

We require a means of representing this information in the same geometric space we are using to describe concepts. Note, however, that concepts generally do not involve all possible pairings of properties between domains. Thus we desire to capture the information in our observation \( x \) that is relevant to a given concept \( C \). To this end, let \( I_C \) denote the indicator set of the properties and property pairings involved in the concept \( C \), i.e., \( C_{ij} > 0 \) for all \((i, j) \in I_C\) and \( C_{ij} = 0 \) for all \((i, j) \notin I_C\). We then define

\[
\Phi_{nm}(C) = \max_{j \in D_n, k \in D_m} \min(s_j, s_k) 
\]

as the largest of the minima of the pairwise similarity values of the observation between pairs of properties in two domains, with

\[
\Phi_{nn}(C) = \max_{j \in D_n} s_j 
\]

Clearly we can use other pairwise aggregation functions in (5) besides the minimum operator, such as products, ordered weighted averages [5] and Minkowski means, but the minimum function is both conservative and computationally attractive.

Thus \( \Phi_{nn}(C) \) is the largest intra-domain similarity of observation \( x \) to a property associated with \( C \) in domain \( D_n \), while \( \Phi_{nm}(C) \) is the largest inter-domain similarity of observation \( x \) to a pairs of properties between domains \( D_n \) and \( D_m \) associated with \( C \), where the individual pairwise similarity is taken as the minimum similarity of the pair. \( \Phi \) is a symmetric, \( M \times M \) dimensional square matrix that captures the best match of observation \( x \) to individual properties within domains and to pairs of properties across domains associated with \( C \), where \( M \) is the number of domains involved in the concept.

We now specify a “connection” matrix \( \Psi(C) \) for the observation \( x \) that is suitable for similarity calculations with respect to concept connection matrices as follows:

\[
\Psi_{ij}(C) = \begin{cases} 
\Phi_{nm}(C) & \text{for all } i \in D_n, \ j \in D_m, (i, j) \in I_C \\
0 & \text{otherwise}
\end{cases} 
\]

where \( i \) ranges over all indices assigned to properties in \( D_n \) and \( j \) ranges over all indices assigned to properties in \( D_m \) that are associated with \( C \), using the same indexing as with \( C \).

Provided that these indices are assigned in consecutive groups to the properties of each domain, the matrix \( \Psi(C) \) will have a block-rectangular structure in which all elements involving allowable property matches between a given pair of domains (i.e., contained in \( I_C \)) will have identical values, equal to the largest allowable pairwise property similarity that exists between the pair of domains. Thus for every allowable pair of properties between two given domains, we credit to all such pairs the highest property pair similarity score for the object between these two domains. Note that this highest score may correspond to a pair of properties having a low but non-zero correlation in the concept, which is perfectly acceptable given that the concept admits such pairings.

Since the matrix \( \Psi(C) \) is of the same dimension as the concept connection matrix \( C \), we can perform a transformation to convert \( \Psi(C) \) into a vector \( \psi(C) \) in the same unit hypercube in which the vector \( c \) resides. The usefulness of this representational scheme will become apparent when we describe the measure to be used for concept/observation similarity.

4.2 Fuzzy subsethood

As distinct from traditional set theory, every fuzzy set is a subset (to a quantifiable fuzzy degree) of every other fuzzy set. The basic measure of the degree to which fuzzy set \( A \) is a subset of fuzzy set \( B \) is fuzzy subsethood, defined as [6]:

\[
S(A, B) = 1 - \frac{M(A \cap B^*)}{M(A)} = 1 - \frac{\sum \max(0, A_i - B_i^*)}{\sum A_i} 
\]

\[
= \sum \min(A_i, B_i^*) 
\]

where \( M(A \cap B^*) \) is the Hamming distance between \( A \) and \( B^* \), the nearest point to \( A \) contained within \( F(2^B) \), the fuzzy power set of \( B \) (i.e., \( F(2^B) \) is the set of all proper fuzzy sets of \( B \)), and \( M(A) \) is the Hamming norm of fuzzy set \( A \). From (8) we have \( 0 \leq S(A, B) \leq 1 \). \( S(A, B) = 0 \Rightarrow A \cap B = \Phi \) (the null fuzzy set), \( S(A, B) = 1 \Rightarrow A \subseteq B \) (\( A \) is a proper subset of \( B \)), and \( 0 \leq S(A, B) < 1 \) implies that \( A \) lies outside of \( F(2^B) \). Note that fuzzy subsethood in general is not symmetric, i.e., \( S(A, B) \neq S(B, A) \).
Figure 2 illustrates, for three different fuzzy sets $A_1$, $A_2$ and $A_3$, their corresponding nearest fuzzy sets $B_{1}^{*}$, $B_{2}^{*}$ and $B_{3}^{*}$ in the fuzzy power set $F(2^B)$ of a fourth fuzzy set $B$. The Hamming distances between these respective sets represents the numerator term in the first line of (8).

Finally, if $0 < S(a,b) < 1$, then $A$ and $B$ have some (non-zero strength) edges in common, but there is no strict subordination of edge connection strengths between them, i.e., $A \cap B \neq A \cup B$.

4.3 Concept/observation similarity

With the above mathematical apparatus, we can now define the similarity $\sigma(C,x)$ between an observation $x$ and a concept $C$ to be the fuzzy subsethood of the corresponding hypercube concept vector $\psi(C)$ in the hypercube observation vector $\psi(C)$:

$$\sigma(C,x) = S(c, \psi(C)) = \frac{\sum_{i \in I_C} c_i \min(x, y_i)}{\sum_{i \in I_C} c_i}$$

Intuitively, this measures the degree to which an observation’s pairwise property similarities relative to $C$, as defined by (5) and (7), equal or exceed those required by the concept. This measure is not only mathematically well-founded, but is computationally appealing as well. Given a library of concepts, one can thus compare an observation to members of this library (making use of cluster-tree based search strategies [7] if desired to reduce the number of comparisons needed) and compile a ranked list of concepts in order of their similarity to the observation. Given the inherently normalized [0, 1] aspect of the fuzzy subsethood measure independent of the number of dimensions, these similarity values can be thresholded to identify concepts to which the observation has adequate similarity. If an observation proves dissimilar (i.e., below threshold) to all concepts in the library, it is classified as a novelty, which warrants further investigation and/or suggests the need for additional concepts.

The form of (11) indicates that our decision measure is a fuzzy nonlinear, relative aggregation function, i.e., a nonlinear combination of the pairwise observation property similarities relative to the concept correlations. This contrasts with other reasoning approaches such as possibilistic logic [8], which represents knowledge whose truth or falsity is uncertain but whose content is precise. Our approach may be considered a form of discrete decision making with multiple objectives [9], where each concept represents a set of (weighted) objectives in the form of the pairwise property similarities making up the concept’s matrix $C$. The decision problem is to select from the library of concepts the one whose objectives are best contained (in a fuzzy sense) by the property similarities of an observation.

5 Level 2/3 Fusion Applications

Level 2 fusion is concerned with situation understanding, i.e., the group characterization of a set of objects derived from Level 1 fusion. Level 3 fusion is concerned with threat assessment, i.e., the prediction of an adversary’s intentions, given their observed group character. In both instances, current fusion system implementations for specific application areas often
embodies *ad hoc* approaches that lack any capacity for generalization to other applications.

### 5.1 Level 2 fusion

Stubberud, Shea and Klamer [10] have pointed out the need for a state-based representation of Level 2 objects as a first step towards establishing a common architecture for Level 2 fusion, such as exists for Level 1 fusion. The starting point for this representation is to perform the detection of candidate Level 2 objects via the clustering of Level 1 objects [7,11]. These authors also proposed certain metrics for use in the association of these Level 2 objects to *templates* in order to provide a symbolic characterization of these objects [12]. These metrics include a normalized $\chi^2$ residual for kinematic distance, cardinality and gap metrics for group compositional differences, a Hausdorff metric for group separation, an invariant moment metric for group formation differences, and an area metric for group region of influence.

We suggest here that the conceptual spaces approach outlined in this paper provides precisely the type of general architecture sought after in [10], but in a richer descriptive context than would be offered by a template representation. In particular, we cite the following matches between our approach and the features sought after in [10]. The different vector components of the Level 2 state prescribed in [10] reside in individual domains of the conceptual space representing an application. The metrics proposed in [12] are associated with properties in these domains. Additional domains, properties and metrics may be added as needed further to characterize Level 2 objects. Rather than describing different regions of the state space symbolically and using rule-based associations of Level 2 objects to templates, we employ the more powerful geometric descriptions of properties and concepts inherent in the conceptual space paradigm. Finally, we supply the additional analytical metrics needed to compare Level 2 objects to concepts.

### 5.2 A Level 2 fusion example

Consider the following Level 2 fusion example, where our objective is to distinguish between “normal” and “aberrant” behavior of a platform. Suppose that we have identified four domains for the conceptual space as described above in Figure 1. Two of them each have two distinct properties that characterize normal behavior, while the remaining two domains each have one property indicative of normal behavior. Bear in mind that each domain in this example may possess an arbitrary number of dimensions.

Let the concept matrix $C$ for normal behavior be given as in Figure 1. The matrix entries identify the typical correlations of property pairs under normal behavior.

Suppose that our observations of a situation have the following vector of similarities with respect to the six properties: $(0.8\ 0\ 0.8\ 0\ 0.8\ 0.7)$. The observation matrix $\Psi$ is then given by (from (5) and (6)):

$$
\begin{bmatrix}
0.8 & 0 & 0.8 & 0 & 0.8 & 0.7 \\
0 & 0.8 & 0.8 & 0.8 & 0.8 & 0.7 \\
0.8 & 0.8 & 1 & 0 & 0.9 & 0.7 \\
0 & 0.8 & 0 & 1 & 0.9 & 0.7 \\
0.8 & 0.8 & 0.9 & 0.9 & 0.9 & 0.7 \\
0.7 & 0.7 & 0.7 & 0.7 & 0.7 & 0.7
\end{bmatrix}
$$

Transforming both $C$ and $\Psi$ into their corresponding hypercube vector points $c$ and $\psi$, and assuming equal salience weights for all properties, we calculate the subsethood of $c$ in $\Psi$ to be 0.884, indicating that the situation is exhibiting reasonably normal behavior. If the similarity of the observation to the third property is reduced from 1 to 0.9, the subsethood of $c$ in $\Psi$ is reduced only slightly to 0.874.

On the other hand, suppose that the observations of the situation have the following significantly reduced similarities to the above six properties: $(0.2\ 0\ 0\ 0\ 0.9\ 0.1)$. This results in the following observation matrix:

$$
\begin{bmatrix}
0.2 & 0 & 0 & 0 & 0.2 & 0.1 \\
0 & 0.2 & 0 & 0 & 0.2 & 0.1 \\
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 \\
0.2 & 0.2 & 0 & 0 & 0.9 & 0.1 \\
0.1 & 0.1 & 0 & 0 & 0.1 & 0.1
\end{bmatrix}
$$

Repeating the subsethood calculation using the corresponding $\Psi$ vector resulting from (13), we obtain a value of only 0.141, thus clearly indicating aberrant behavior of the platform. Note that all values of subsethood will be modified in accordance with (9) in the case where the property pair salience weights $\omega_k$ are not all equal.

### 5.3 Level 3 fusion

The conceptual space approach can be applied to Level 3 fusion as well. The prediction of the intentions of an adversary can be approached by at least two different methods, both of which are amenable to operating on a conceptual space construct. The most straightforward of these is to define concepts pertaining to predicted *intentions*, and then calculate the subsethood of these concepts in the Level 2 situational concepts determined from observations, as described in Section 4 above. This subsethood value provides a direct measure of the degree to which an intention concept is matched by observations. This approach subsumes both rule-based and pattern-based intention concepts.

The other approach employs game theory to explore various courses of action, and makes its predictions based upon the most favorable strategies in which an adversary could engage. Here, the space of the game can be the conceptual space, and the menu of candidate moves can
be embodied as concepts. For example, the concept “attack position \((x, y)\)” specifies a set of kinematic, composition, formation and area of influence properties that a Level 2 object must possess in order to carry out such an attack. The degree to which an observation matches to this concept is thus predictive of possible enemy intentions.

6 Conclusion

In this paper, we have extended Gardenfors’ notion of a “concept” in a conceptual space to be represented as a point in a unit hypercube that captures all of the concepts’ properties, salience weights and correlations. This description of a concept can be learned from a training data set. We also present a method for describing a particular observation as a point in the same unit hypercube. We propose the fuzzy subsethood metric for measuring the similarity of an observation to a concept. This metric yields inherently normalized values in \([0,1]\), which allows concept similarities to an observation to be ranked and thresholded without further normalization in order to declare situations and/or alert to novelties. Finally, we discuss the application of this technology to Level 2 and Level 3 data fusion problems and provide an example illustrating the detection of anomalous behavior.

7 References