A Metric Conceptual Space Algebra

Benjamin Adams and Martin Raubal

Departments of Computer Science and Geography
University of California, Santa Barbara
badams@cs.ucsb.edu, raubal@geog.ucsb.edu

Abstract. The modeling of concepts from a cognitive perspective is important for designing spatial information systems that interoperate with human users. Concept representations that are built using geometric and topological conceptual space structures are well suited for semantic similarity and concept combination operations. In addition, concepts that are more closely grounded in the physical world, such as many spatial concepts, have a natural fit with the geometric structure of conceptual spaces. Despite these apparent advantages, conceptual spaces are underutilized because existing formalizations of conceptual space theory have focused on individual aspects of the theory rather than the creation of a comprehensive algebra. In this paper we present a metric conceptual space algebra that is designed to facilitate the creation of conceptual space knowledge bases and inferencing systems. Conceptual regions are represented as convex polytopes and context is built in as a fundamental element. We demonstrate the applicability of the algebra to spatial information systems with a proof-of-concept application.

1 Introduction

In recent years there has been an increasing demand for research on the representation and modeling of cognitive phenomena for spatial information systems [21, 22]. Semantic similarity measurement in particular has been an active area of research for spatial applications [17, 29]. Since human users interface and interoperate with these systems, they must have a means for representing the conceptual structures that exist in the users' minds, especially those concepts that are related to spatial cognition. Although geometric modes of concept representation have not been as widely adopted as other representational frameworks for cognitive modeling, they have garnered interest from researchers in the spatial sciences, because many spatial concepts are intrinsically thought of in terms of their geometric and topological features.

Models of human cognitive processes require a formal representation that a computer system can interpret. The two prevailing frameworks for representing cognitive processes are the symbolic and connectionist methods [9]. The symbolic method aims at modeling high-level abstract concepts using symbol manipulation schemes. Inferences are often the result of first-order logical operations on the symbols in the model. The connectionist method attempts to model cognition in a way that more closely compares to the biological neural structure of the brain, mathematically represented as nodes and their weighted connections.

While useful for many cognitive computational tasks, both of these representational frameworks do not perform well at modeling certain aspects of cognition, specifically semantic similarity and concept combination. Semantic similarity measurement is fundamental to the task of cognitive categorization, because conceptual units are classified with other conceptual units with which they are most similar [28]. Using symbolic representations, the combination of concepts is most often measured as the set-theoretic intersection of the properties of two classes, but this method fails when combining concepts without shared properties [10]. For example, *stuffed gorilla* combines the concepts *stuffed* and *gorilla*. However, the *gorilla* concept has no intersecting properties with the *stuffed* concept because it is a living thing. In addition, ad hoc concepts, which by definition are concepts that combine from different domains, are similarly difficult to represent using symbolic representations [16]. In the case of connectionist representations, even small connectionist models can be highly complex and become unwieldy for representing semantics on the level needed for operating with concepts.

As a complement to the two representational frameworks listed above, Gärdenfors has introduced conceptual space theory [9]. Conceptual spaces are geometric and topological structures that represent concepts as convex regions in multi-dimensional domains. This theory constitutes a mid-level spatialization approach to concept representation and is particularly suited as a framework for spatial information systems. Conceptual spaces can model semantic similarity naturally as a function of distance within a geometric space, and conceptual regions are subject to geometric operations such as projections and transformations that result in new concept formations.

Critics of conceptual space theory have contended that its usefulness has only been demonstrated for simplistic cases with little abstraction and using formalizations that are designed for specific contexts [33]. It is our position that rather than being due to theoretical limitations, the difficulty in assessing the experimental worth of conceptual spaces has been in part that no conceptual space algebra exists with well-defined operations that allow one to build and reason with complex conceptual space structures. To help rectify that situation, in this paper we present a metric conceptual space algebra, consisting of formal definitions of its components and operations that can be applied to them. The work builds upon previous formalizations of conceptual spaces but aims to be more comprehensive both as a mathematical model and as a launching pad for computational implementation. Our key contributions are the formalizations of query operations for semantic similarity measurement and concept combination.

Section 2 introduces conceptual space theory and previous formalization approaches. In section 3, we define a conceptual space algebra with its components. Concepts are thereby represented as convex polytopes. In addition, contrast classes and context are formally defined. Section 4 presents the algebraic operations, i.e., core metric operations, and query operations for similarity and concept combination. Section 5 applies the conceptual space algebra to the problem of comparing countries and regions of the world with different contexts. The final section presents conclusions and directions for future research.

2 Related Work

The theory of conceptual spaces was introduced as a framework for representing information at the conceptual level [9]. Conceptual spaces are based on the paradigm of cognitive semantics, which emphasizes that meanings are mental entities, i.e., mappings from symbols to conceptual structures, which refer to the real world [20]. They can be utilized for knowledge representation and sharing, and account for the fact that concepts are dynamic and change over time [4, 26].

A conceptual space is a set of quality dimensions with a geometric or topological structure for one or more domains. Domains are represented by sets of integral dimensions, which are distinguishable from all other dimensions, e.g., the color domain. Concepts cover multiple domains and are modeled as n-dimensional regions. Every instance of the corresponding category is represented as a point in the conceptual space. This allows for expressing the similarity between two instances as a function of the spatial distance between their points. Recent work has investigated the representation of actions and functional properties in conceptual spaces [12].

Vector algebra offers a natural framework for representing conceptual spaces. A conceptual vector space can be formally defined as $C^n = \{(c_1, c_2, \ldots, c_n) | c_i \in C\}$ where the c_i are the quality dimensions [25]. Vector spaces have a metric and therefore allow for the calculation of distances between points in the space. This can also be utilized for measuring distances between concepts, either based on their approximation by prototypical points or regions [30]. Calculating semantic distances between instances of concepts requires that all quality dimensions of the space must be represented in the same relative unit of measurement. Given a normal distribution, this can be achieved by applying the z-transformation for these values [7]. Different contexts can be represented by assigning weights to the quality dimensions of a conceptual vector space. C^n is then defined as $\{(w_1c_1, w_2c_2, \ldots, w_nc_n) | c_i \in C, w_j \in W\}$ where W is the set of real numbers. The use of convex hull and Voronoi tessellation algorithms can be used to learn conceptual space regions from a set of data points [13].

Work has been done to link conceptual space theory to established representational frameworks. Conceptual spaces are mid-level representations and they have been bridged to higher-level symbol representations [3]. The geometric representation of concepts has been extended to a fuzzy graph representation as well [27]. However, the work done so far has not provided an integrated framework that encompasses the full suite of conceptual space principles within a mathematically defined geometric and topological structure, which is the aim of the algebra presented here.

3 Formal Definitions

In this section we present a formal definition of a metric conceptual space and its components. The conceptual space definition is mathematical and designed for the practical goal of facilitating the construction of conceptual space knowledge bases. For this reason, the convex regions used to represent concepts are specified with more explicitness than in previous formalizations of conceptual spaces. Convex regions are defined as a convex polytopes [14], which are generalizations of polygons to n dimensions. The $n-1$ dimensional faces of a convex polytope are called facets. This definition of a concept region was chosen because operations on polytopes are computationally tractable for domains composed of more than two dimensions. Curved regions can be approximated using polytopes in much the same way that polygon primitives are used to describe more complex structures in geographic information systems (GIS) and computer graphics applications. In addition, there are mathematical representations for convex polytopes that are generalizable over any number of dimensions. This is important, because unlike in GIS and graphics applications, the number of dimensions in a domain can be arbitrarily large.

A designation of the context is required for many conceptual space algebra operations. Methodologies for representing context for similarity measurement has been an active research area [19]. We extend the notion of context for similarity as weights on domains as well as quality dimensions. Take, for example, a conceptual space with a *color* domain that is composed of three quality dimensions: *hue*, *value*, and *saturation*. It is conceivable that one may want to weight the entire *color* domain lower in a *night* context [36], while also weighting *value* higher than *hue* and *saturation*. This secondary weighting has the effect of making a dark red color more similar to a dark blue color than to a light red color.

The role of context is not confined to similarity measurement. When combining concepts the salience of the domains for each concept helps to determine which regions override other regions. Given the ubiquity of context, we define a context as a set of salience weights that can be applied to components of any type in the conceptual space. This definition leaves open the option of applying salience to objects in the conceptual space in a manner beyond what is discussed in this paper, which will facilitate extending the operation set of the algebra. For example, one can create a context for a set of instances where each instance in the set is given its own salience weight for prototype learning.

3.1 Metric Conceptual Space Structure

A metric conceptual space is a multi-leveled structure. A distinction is made between the representation of the geometric elements (regions and points) and the conceptual elements (concepts, properties, and instances). In contrast to other formalizations of conceptual spaces, regions and points are associated with only one domain each, and not with the conceptual space as a whole. Concepts and instances, on the other hand, span across one or more domains. This structure facilitates semantic similarity measurements for concepts and instances that take into account different distance measurements for within and between domains as well as concept combination operations that operate domain-by-domain.

The following definitions are organized in a top-down way beginning with the definition of a conceptual space and defining each component of this space in turn. We refer to concepts, properties, and instances as *objects* in the conceptual space.

Definition 1. *A* metric conceptual space *is defined as a 6-tuple,* $S = \langle \Delta, \Gamma, \breve{I}$, ⬗, $K, c\rangle$.

- Δ *is a finite set of* domains*, where a* domain $\delta \in \Delta$.
- Γ *is a finite set of* concepts*, where a* concept $\gamma \in \Gamma$.
- $\breve{I}$ *is a finite set of* instances*, where an* instance $\breve{\imath} \in \breve{I}$.
- ⬗ *is a finite set of* contrast classes*, where a* contrast class ⬗ $\in$ ⬗.
- K *is a finite set of* contexts*, where a* context $k \in K$.
- c *is a constant* similarity sensitivity *parameter.*

The set components of a conceptual space are dynamic and can be modified by applying algebraic operations. The conceptual space algebra defines a number of operations that take elements from one or more of the components of the conceptual space and produce values. In some cases query operations will produce numeric or Boolean values, but in other cases they will produce higher-level structures such as new concepts and modify the existing set. In the latter case, the products are inserted into the appropriate set component. For example, an operation to learn a new concept will add the new concept into the Γ component of the conceptual space.

3.2 Domains and Quality Dimensions

Definition 2. *A* domain *is defined as a set of quality dimensions,* $\delta = Q$. Q *is the finite set of integral* quality dimensions *that form the* domain*, where a* quality dimension $q \in Q$. $\forall q, q \in \delta \wedge \delta \neq \delta' \Rightarrow q \notin \delta'$.

Definition 3. *A* quality dimension *is defined as a triple,* $q = \langle \hat{\mu}, \hat{r}, \hat{o} \rangle$.

- $\hat{\mu}$ *indicates the* measurement level *or scale of the dimension, where* $\hat{\mu} \in \{ratio, interval, ordinal\}$.
- $\hat{r}$ *indicates the* range *of the dimension, where* $\hat{r}$ *is a pair* $\hat{r} = \langle min, max \rangle$.
- $\hat{o}$ *indicates whether the dimension is* circular*, where* $\hat{o} \in \{true, false\}$.

The quality dimensions in a conceptual space represent a means for measuring and ordering different quality values of objects in the space (in the case of concepts these values might be a range of values). There are four widely-recognized scales of measurement – nominal, ordinal, interval, and ratio – that can be used to assign values to data, and each of these measurement levels has associated with it different mathematical properties [32]. The $\hat{\mu}$ component of a quality dimension can specify the quality dimension scale as ordinal, interval, or ratio. Interval and ratio scales both work naturally for quality dimensions because differences in measurements can be easily compared due to the fact that the units for these scales are equalized. An ordinal scale's values are rank ordered and are consistent with the ordering operations of a conceptual space algebra. However, conclusions of semantic similarity for ordinal quality dimensions should be made with care, taking into account the fact that ordinal scales do not have equalized scales. Psychometric analysis techniques for imposing a distance measurement on an ordinal scale (e.g., Rasch models) can be used to convert an ordinal scale to interval scale [24]. Since, we are primarily interested in quality dimensions as ways of specifying measurement and for use in ordering operations, nominally scaled quality dimensions are not directly supported by this definition. However, it is possible to represent the values on a nominal scale as properties in a domain. One approach for modeling geographic nominal data values as regions in a conceptual space has been shown in [2].

Quality dimensions can be phenomenal or scientific, which means they represent subjective psychological dimensions or are defined by positivist scientific theories, respectively [11]. Conceptual spaces are equally capable of representing both types of dimensions and there is no distinction made between scientific and phenomenal dimensions in the formal definition. A circular dimension is one that wraps around, so the maximum distance value is $\frac{\hat{r}_{max}-\hat{r}_{min}}{2}$. For example, the hue dimension in the color domain is a circular dimension with value range of $[0, 2\pi]$, and any measured distance will be $< \pi$.

3.3 Concepts, Properties, and Instances

Definition 4. *A* concept *is defined as a pair,* $\gamma = \langle \Diamond, P \rangle$.

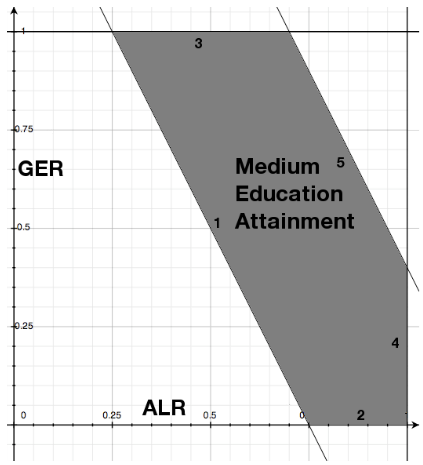
- $\Diamond$ *is a finite set of* convex regions*, where there is an injective relation between* $\Diamond$ *and* Δ*. That is, there is a one-to-one relationship from regions in the set to domains and there can only be one region per domain.*
- P *is a* prototypical instance.

Definition 5. *A* property *is defined as a* concept *with* $|\Diamond| = 1$.

A concept is a collection of convex regions across one or more domains and an associated prototypical instance. $\forall p, p \in P \Rightarrow \exists \diamond, \diamond \in \Diamond \wedge p \in \diamond$. The prototypes or representative members of a concept play an important role in categorization [28]. There is experimental evidence that the perceived similarity of an object to a prototypical exemplar is used by humans during classification [15]. Given the prototypical instance(s) of one or more concepts, one can derive the regions that compose it using a Voronoi tessellation technique [13]. Conversely, a prototypical instance can be identified by finding the point of central tendency for a set of exemplar points in each domain. The measurement level of the quality dimensions determines how the central tendency is calculated: the geometric mean (or the arithmetic mean of the natural logarithm scale) for ratio scaled, the arithmetic mean for interval scaled, and the median for ordinal scaled dimensions.

Definition 6. *A* convex region $\diamond$ *is defined as a convex polytope in the n-dimensional space corresponding to a given domain,* δ.

As defined, a convex region in a conceptual space can be represented as either 1) a set of vertices that constitute the convex hull of the region or 2) a bounded



$$\begin{array}{c} \\ \mathbf{1} \\ \mathbf{2} \\ \mathbf{3} \\ \mathbf{4} \\ \mathbf{5} \end{array} \begin{array}{c} \begin{array}{cc} \textbf{ALR} & \textbf{GER} \end{array} \\ \begin{pmatrix} -0.894 & -0.447 \\ 0.0 & -1.0 \\ 0.0 & 1.0 \\ 1.0 & 0.0 \\ 0.894 & 0.447 \end{pmatrix} \end{array} \begin{pmatrix} q_1 \\ q_2 \end{pmatrix} \leq \begin{pmatrix} -0.671 \\ 0.0 \\ 1.0 \\ 1.0 \\ 1.073 \end{pmatrix}$$

Fig. 1: Example region in education index domain and H-Polytope representation

intersection of half-spaces that can be written as a set of linear inequalities [14]. The first representation is called a V-polytope and the second representation a H-polytope. The H-polytope and V-polytope representations are equivalent, but some important core operations, such as point inclusion, have much better combinatorial complexity when starting with a H-polytope representation.

A region within an n-dimensional domain can be written in the H-polytope matrix form $Aq \leq b$, where q is a variable transpose vector and each q_i corresponds with an $\in Q$:

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{n1} \\ a_{21} & a_{22} & \cdots & a_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{nm} \end{pmatrix} \begin{pmatrix} q_1 \\ q_2 \\ \vdots \\ q_n \end{pmatrix} \leq \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}$$

The values in any row i of the A matrix and b transpose vector correspond to the coefficients of the linear inequality that defines the i^{th} half-space boundary of the polytope.

Figure 1 shows a region with five facets in a domain built from two dimensions: adult literacy rate (ALR) and gross enrollment ratio (GER). The domain is based on the United Nations' measure of educational attainment [35]. A country's education index is equal to $\frac{2}{3} \times ALR + \frac{1}{3} \times GER$. Here we define a concept of medium education attainment as the region where $0.5 \leq$ education index ≤ 0.8.

Definition 7. *An* instance $\breve{\imath}$ *is defined as a finite set of points with an injective relation to Δ. That is, there is a one-to-one relationship from points in the set to domains and there can only be one point per domain.*

Instances, which can be thought of as real-world objects or data points for training sets, are represented by a set of points (or vectors) in one or more domains.

Definition 8. *A* point *p is defined as a vector of quality dimension values in the n-dimensional space corresponding to a given domain, δ.*

3.4 Contrast Class

Definition 9. *A* contrast class ⬗ *is defined as a region in a unit hypercube that corresponds to a domain in the conceptual space. Each dimension of the hypercube corresponds to one quality dimension in the domain. The region is specified by one or two parallel hyperplanes that intersect the unit hypercube:*

$$min : -a_1x_1 - a_2x_2 - \cdots - a_nx_n \leq -b_1$$
$$max : a_1x_1 + a_2x_2 + \cdots + a_nx_n \leq b_2$$

Contrast classes are described as a special type of property (or class of properties) by Gärdenfors. We define it as a unique type of element for the algebra, because it is used differently from a normal property by algebraic operations. It is more appropriately thought of as a function that describes sub-regions relative to regions in a domain. In a one-dimensional domain the contrast class region is bounded by two points on a unit line segment; in two dimensions the region is bounded by two parallel lines intersecting a unit square; and so on. The operation for combining a contrast class with a concept is detailed in section 4.

Some common contrast classes are *large, small, old, young, northern, southern*. The example shown in figure 1 is the result of combining a *medium* contrast class to the *education attainment* property. *Education attainment* covers the square area [0,1] on both dimensions, and the *medium* contrast class is projected on that region resulting in the *medium education attainment* sub-region.

3.5 Context

Definition 10. *A* context *k is defined as a finite set of salience weight, conceptual space component pairs $k = \langle \omega, component \rangle$. The conceptual space components in the context must all be of the same type (e.g., quality dimensions), which is referred to as the* context type. *In addition, $\sum_{i=1}^{n} \omega_i = 1$ where each weight ω_i in a context has a value $0 \leq \omega_i \leq 1$.*

If a context is used in an operation that is applied to a context typed conceptual space component that is not included in the context then the salience weight of that component is 0. For example, if a similarity operation is applied to two instances that span the same three domains and the context only contains weights for two of the three domains then the third dimension's salience weight will be 0 for the operation.

4 Algebraic Operations

In this section we introduce the algebraic operations that can be placed on the components of a conceptual space. We organize the operations into core metric operations on points and regions followed by similarity and concept combination query operations. We use shorthand notations for the components (Table 1).

4.1 Core operations

Regions and points are the most primitive objects in a conceptual space, each existing within a single $\delta \in S$. Since a domain is a metric and topological space, all of the operations that can be applied to regions and points in a topological vector space are applicable. The computational implementation of these operations essentially reduces to problems of numeric linear algebra, which are well-established. With the understanding that there are many more operations possible on these primitive types, we present the particulars of how the H-polytope representation of conceptual regions can be used to implement intersection and inclusion operations algorithmically. In addition, the within domain distance metric for points and regions is defined.

$\boldsymbol{\diamond_1 \bigcap \diamond_2 \rightarrow \diamond_{new}}$**.** This function calculates the intersection of two convex regions. The result is a new convex region or an empty set. The first step of an algorithm to calculate the intersection of two H-polytopes is to get the union of the two systems of linear inequalities. It is possible that this union will result in redundant inequalities, so a linear programming problem is constructed to remove these redundancies: Given two regions $\diamond_1$ and $\diamond_2$ represented as $Aq \leq b$ and $Sq \leq t$, respectively, maximize each inequality $s^T q$ in $\diamond_2$ subject to $Aq \leq b$ and only add the inequality to the union if the optimal value is less than or equal to t [8]. There are several algorithmic techniques used for solving linear programming problems, the most popular being the Simplex method [6]. The Simplex method performs very well in most cases, averaging a number of iterations that is less than three times the number of inequalities in the set [23]. The total computational complexity of the intersection operation is therefore linear with respect to the number of inequalities in most cases.

$\boldsymbol{p \in \diamond \rightarrow}$ ***Boolean.*** The inclusion operation given a point and a region represented as a H-polytope is equivalent to testing whether the point satisfies the

Table 1: Notation for named elements

Notation	Meaning	Example
$\gamma^{concept}$	named concept	$\gamma^{european\ state}$
$\diamond^{concept}_{domain}$	region of concept in domain	$\gamma^{european\ state}_{coordinates}$
δ^{domain}	named domain	$\delta^{coordinates}$
$\delta^{domain}_{quality\ dim}$	quality dimension in domain	$\delta^{coordinates}_{latitude}$
$\blacklozenge^{domain}_{contrast\ class}$	contrast class in domain	$\blacklozenge^{coordinates}_{southern}$
$\delta\left(\gamma_P\right)$	domain of property	$\delta\left(\gamma^{temperate\ zone}\right) = \delta^{coordinates}$
$Q\left(\delta^{domain}\right)$	quality dimensions of domain	$Q\left(\delta^{coordinates}\right) = \{longitude, latitude\}$

entire system of linear inequalities. The computational complexity is linear with respect to the number of facets in the polytope.

***dist*(p_1, p_2, k) $\rightarrow \mathbb{R}$.** The distance metric for two points that exist within the same domain is defined as a weighted Euclidean metric, where the weights are determined by a quality dimension-typed context, k.

$$dist(p_1, p_2, k) = \sqrt{\sum_{i=1}^{n} \omega_i (p_{1_i} - p_{2_i})^2}$$

where n is $|Q|$, i is an index to an ordering of Q, q_i is the i^{th} element in that ordered set, and $(\omega_i, q_i) \in k$. If a dimension is circular then the difference $p_{1_i} - p_{2_i}$ in the above equation will be modulo the range of the quality dimensions divided by two. Euclidean distance measure, as opposed to another instance of the Minkowski metric, was chosen because of experimental results that show its cognitive plausibility for measuring the similarity of concepts composed of integral qualities [9, 18]. When the qualities are separable, the city-block distance was found to be more appropriate, which is captured by the similarity operation.

Normalization of dimensions is important to ensure that a change in units does not result in a different distance measurement and subsequently a different similarity measurement. However, given the variety of options depending on the structure of domains, we reserve normalization as a preprocessing operation rather than an integral component of the distance measure.

4.2 Query operations

Similarity. Experiments on similarity cognition have shown that the similarity of two objects can be measured as an exponentially decaying function of the distance between the two objects: $sim(d) = e^{-cd}$ [31]. The following similarity operation utilizes a compound distance function that takes into account the structural distinction of separable and integral dimensions.

$sim(\breve{\imath}_A, \breve{\imath}_B, k, K) \rightarrow \mathbb{R}$. Given two instances, $\breve{\imath}_A$ and $\breve{\imath}_B$, a domain-type context, k, and a set of quality dimension-type contexts, K, this function calculates a distance between $\breve{\imath}_A$ and $\breve{\imath}_B$. Let $\Delta_{\breve{\imath}} = \Delta(\breve{\imath}_A) \bigcap \Delta(\breve{\imath}_B)$. The distance between two instances is a weighted sum of all of the within domain Euclidean distance measures for each $p \in \breve{\imath}$:

$$d(\breve{\imath}_A, \breve{\imath}_B, k, K) = \sum_{j=1}^{|\Delta_{\breve{\imath}}|} k_j \times \sqrt{|Q(\delta_j)|} \times dist(p_j(\breve{\imath}_A), p_j(\breve{\imath}_B), K_j)$$

where j is an index to an ordering of $\Delta_{\breve{\imath}}$. Ignoring context weights, the result of this distance function is a composite value that is $\geq$ the Euclidean distance and $\leq$ the city-block distance, if all the quality dimensions were in one multi-dimensional space. The context and context set parameters allow one to apply saliences on both domain and quality dimension levels. Because Euclidean and city-block metrics are being mixed, each weighted within-domain Euclidean distance measure is also normalized by the square root of the cardinality of the domain to prevent low dimensional domains from having more salience than high dimensional domains. Following this distance function we get the following similarity function for two instances:

$$sim(\breve{\imath}_A, \breve{\imath}_B, k, K) = e^{-cd(\breve{\imath}_A, \breve{\imath}_B, k, K)}$$

$sim(\gamma_A, \gamma_B, k, K) \rightarrow \mathbb{R}$. There are many possible methods for measuring the distance between two regions within a domain. The simplest method is just to measure it as the distance between prototypical points within each region. The analogue to that method for a similarity measure between concepts would be to measure the distance between prototypical *instances* for the two concepts using the similarity function just described. Other methods have been proposed to measure distance between spatial conceptual regions using the vertices of the convex hull of the region [30]. The advantage to these methods is that they allow for asymmetrical distance measurements, the lack of which is a common criticism of geometric models of similarity [34]. Using the V-polytope form of the regions, the methods can be used to calculate the within-domain distance for each domain, which can then be summed in a weighted form as above. The distance from an instance to a concept can also be calculated.

Concept combination. Gärdenfors describes techniques for combining concepts in conceptual spaces, but his methodology has not been formalized yet [9]. Here we describe three concept combination operations using the components of a conceptual space as defined above. The operations are property-concept, concept-concept, and contrast class-concept combinations. For these operations it is important to note that one concept is the *modifier* concept and the other is the *modified* concept. This distinction is linked with the importance of the ordering of concepts in linguistic expressions. For example, the concept combination *green village* is distinct from the concept combination *village green*. We follow a convention that the *modifier* concept is the first parameter and the *modified* concept is the second parameter of any combination operation.

The following algorithms describe how the regions of concept combinations are formed. With all three concept combination operations not only new regions but also new prototypical points need to be learned. The process is the same for all three. For any newly created region $\diamond_{new}$, the new prototypical point is set equal to the centroid of $\diamond_{new}$. Alternately, in the case that an associated instance set is available, the prototypical point can be learned from the set of instances $\in \diamond_{new}$.

$combine\,(\gamma_P, \gamma_C) \rightarrow \gamma_{new}$. Algorithm 1. The combination of a property and a concept is the simplest case. There is no need to specify the salience of the domains, because it is understood that $\delta(\gamma_P)$ is of higher salience. In the case that γ_C does not have a region specified for the domain of γ_P, the property's region is added to the concept. When the property region is part of a specified domain for γ_C then there are two possible outcomes. The property region and the concept region for that domain can overlap, in which case the new concept region is the intersection of the regions. If they do not overlap, the property region will override the concept region in that domain. For example, the property-concept combination *purple mountain* will override the *mountain* region in the *color* domain to the *purple* region assuming that the *color* region for *mountain* is in another area of the *color* domain. The other domains are unaffected.

Algorithm 1 Property-Concept Combination

operation combine$(\gamma_P, \gamma_C) \rightarrow \gamma_{new}$
if $\Delta(\gamma_C) \ni \delta(\gamma_P)$ **then**
  if $\diamond^{\gamma_P}_{\delta(\gamma_P)} \bigcap \diamond^{\gamma_C}_{\delta(\gamma_P)} = \emptyset$ **then**
    $\gamma_{new} \Leftarrow \left(\Diamond(\gamma_C) - \left\{\diamond^{\gamma_C}_{\delta(\gamma_P)}\right\}\right) \bigcup \left\{\diamond^{\gamma_P}_{\delta(\gamma_P)}\right\}$
  else
    $\gamma_{new} \Leftarrow \left(\Diamond(\gamma_C) - \left\{\diamond^{\gamma_C}_{\delta(\gamma_P)}\right\}\right) \bigcup \left\{\diamond^{\gamma_P}_{\delta(\gamma_P)} \bigcap \diamond^{\gamma_C}_{\delta(\gamma_P)}\right\}$
  end if
else
  $\gamma_{new} \Leftarrow \Diamond(\gamma_C) \bigcup \left\{\diamond^{\gamma_P}_{\delta(\gamma_P)}\right\}$
end if
return γ_{new}

combine $(\gamma_A, \gamma_B, K_A, K_B) \rightarrow \gamma_{new}$. Algorithm 2. As with similarity, context plays an important role in concept combinations. When combining two concepts that both span more than one domain, only the regions in a subset of the domains will be affected depending on the context. For both concepts a salience weight is given for each domain (i.e., a domain-type context). If the domains are shared by the two concepts then the context will determine which concept has precedence. Otherwise, the new concept will adopt the region from the concept for which the domain is specified. Currently, the weights are only for comparison, therefore values of 0 and 1 are sufficient. However, room is left for more complex combination operations that take into account the differences in weight values.

combine(⬗, γ) $\rightarrow \gamma_{new}$. The operation to apply a contrast class to a concept only affects the domain for which the contrast class is defined, which we refer to as δ_{CC}. For the sake of brevity, the following is a high-level description of the operation, which at a low-level relies on standard geometric operations. Let $\diamond$ be the region of γ in δ_{CC} and p the prototypical point $\in \diamond$. Find the minimum bounding box around $\diamond$, which gives a range magnitude for each dimension of $\diamond$. Stretch the contrast class unit hypercube (and min, max hyperplanes accordingly) to the size of the minimum bounding box. Center this stretched version over p and intersect the hyperplane(s) with $\diamond$ to get $\diamond_{new}$. γ_{new} is equal to γ

Algorithm 2 Concept-Concept Combination

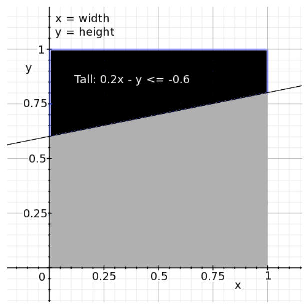
operation combine($\gamma_A, \gamma_B, K_A, K_B$) $\rightarrow \gamma_{new}$
Let $\Delta_{new} = \Delta(\gamma_A) \bigcup \Delta(\gamma_B)$
Let $\gamma_{new} = \emptyset$
for all $\delta \in \Delta_{new}$ **do**
  if $\delta \notin \Delta(\gamma_B)$ **then**
    insert $\diamond_\delta^{\gamma_A}$ into γ_{new}
  else if $\delta \notin \Delta(\gamma_A)$ **then**
    insert $\diamond_\delta^{\gamma_B}$ into γ_{new}
  else { // check context}
    if $K_{A\delta} > K_{B\delta}$ **then**
      $\gamma_{new} = \gamma_{new} \bigcup combine(\gamma_{A\delta}, \gamma_{B\delta})$
    else
      insert $\diamond_\delta^{\gamma_B}$ into γ_{new}
    end if
  end if
end for
return γ_{new}

in all other domains with $\diamond_{new}$. Figure 2 illustrates these steps with a contrast class *tall* and concept *mountain* in a *size* domain with two dimensions: *height* and *width*. This operation can be applied recursively. For example, *tall* can be applied again to *tall mountain* to obtain a region for *very tall mountain*.

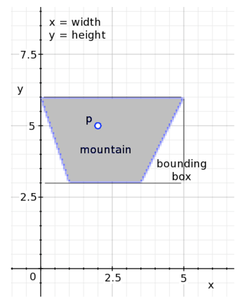
5 Application: Country Concept Comparison

The developed metric conceptual space algebra provides a means for creating complex conceptual space structures and applying similarity and concept combination operations to concepts represented in the space. In order to demonstrate the functionality of these algebraic operations and its use for spatial problems with high dimensional data, we present a case study where the algebra is used for the comparison of countries and regions of the world.

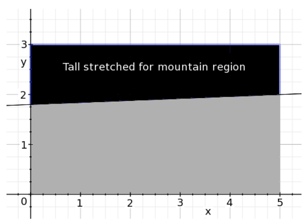
The countries of the world and the groups to which they are classified are complex concepts. The United Nations Development Program (UNDP), for example, divides the countries of the world into eight mutually exclusive classes: Arab States, East Asia and Pacific, Latin America and the Caribbean, South Asia, Southern Europe, Sub-Saharan Africa, Central and Eastern Europe and the CIS, and High-Income OECD [35]. This classification scheme is based on a combination of cultural, geographic, political, historical, and economic factors, but as such it does not afford the ability to make more nuanced comparisons between countries and regions. With a conceptual space representation these factors can be organized into separable domains. The similarity of countries can be compared based on context, and concept combination can generate new classes.



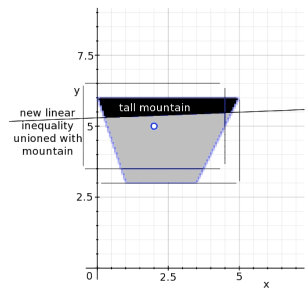
(a) Tall contrast class



(b) Mountain region



(c) Stretched contrast class



(d) Tall mountain region

Fig. 2: Example contrast class-concept combination: tall + mountain

5.1 Data collection

Data for 155 countries were aggregated from the CIA World Factbook, UNDP, and the World Resources Institute [5, 35, 37] and were used to represent each country as an instance in a conceptual space with 16 quality dimensions organized into six domains (Table 2). The UNDP classes were represented as concepts formed by taking the convex hulls of country instances in that class for each domain. We defined the prototypes of these classes as the set of points formed by mean values of all the instances in the class. For the population and size dimensions the values were scaled logarithmically. All data values and dimensions were then normalized so that each quality dimension had a range [0,1].

5.2 Results

For demonstrating the similarity operation, three contexts were created, which we refer to as "natural resources", "geographic", and "human issues"[1]. Table 2 presents the weights for each context. Using the instance similarity operation we

[1] These context weights and their associated labels were chosen merely to illustrate the similarity operation, without making a claim for cognitive plausibility.

Table 2: Country conceptual space domains and contexts

Domains	Quality Dimensions	Context N.R.		Context Geo.		Context H.I.	
Size	*Land area*	0.4	0.8	0.3	0.5	0.1	0.5
	Water area		0.2		0.5		0.5
Coordinates	*Latitude*	0.0	–	0.4	0.7	0.0	–
	Longitude		–		0.3		–
Land type %	*Forests*	0.5	0.5	0.3	0.2	0.1	0.0
	Grasslands		0.0		0.2		0.0
	Wetlands		0.1		0.2		0.0
	Croplands		0.4		0.2		0.5
	Barren		0.0		0.2		0.0
	Urban		0.0		0.0		0.5
Education	*Adult lit. rate*	0.0	–	0.0	–	0.3	0.7
	Gross enrol. ratio		–		–		0.3
Economic	*GDP index*	0.1	1.0	0.0	–	0.2	1.0
Demographic	*Population*	0.0	–	0.0	–	0.3	0.4
	Pop. growth rate		–		–		0.4
	Urban pop. (%)		–		–		0.2

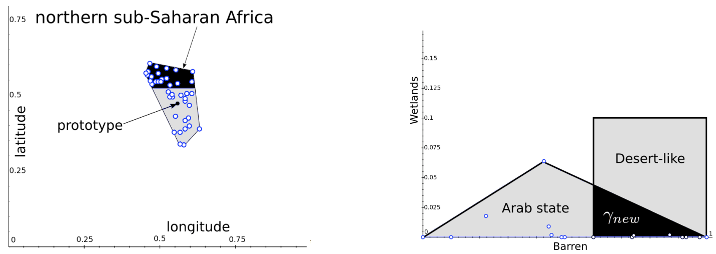
Table 3: Top 5 most similar countries to Turkey by context

Natural Res.	Geography	Human Issues
Thailand 0.877	Kyrgyzstan 0.833	Spain 0.890
Malaysia 0.875	Spain 0.832	Zimbabwe 0.872
Colombia 0.872	Italy 0.831	Uruguay 0.840
Mexico 0.867	Nepal 0.817	Greece 0.827
Viet Nam 0.855	Uzbekistan 0.815	Italy 0.815

found the similarity between every pair of countries for a given context. Table 3 shows a sample of these similarity results for Turkey.

To demonstrate the use of contrast classes we defined a *northern* contrast class for the *coordinates* domain as $\delta^{coordinates}_{latitude} \geq 0.5$ and combined it with the eight UNDP classes. This contrast class corresponds roughly to the top half – along the latitude dimension – of any region (re-centered over the prototype) in the *coordinates* domain. Figure 3a shows the results of combining *northern* with the coordinates region of *sub-Saharan Africa*. Also shown are all the instances with points in the coordinates domain that lie within that region.

Next we created a property region in the land type domain to describe the property *desert-like*. This region is defined as the area where *barren* ≥ 0.6 and *wetlands* ≤ 0.1. The property was combined with the *arab state* concept to create a *desert-like arab state*. Figure 3b shows a two-dimensional projection of the land type domain with the result of this combination.



(a) Northern sub-Saharan Africa in *coordinates*

(b) Desert-like Arab state in 2D *land type* projection

Fig. 3: Concept combinations

Finally, to test complex concept combination we created an ad hoc category for arid, highly educated countries with a large urban population. In the land type domain it was represented by a region where $barren \geq 0.5$, $forests \leq 0.1$, and $wetlands \leq 0.05$; in the education domain it was the region where $\frac{2}{3} \times ALR + \frac{1}{3} \times GER \geq 0.8$; and in the demographic domain it was the region where urban population > 0.8. The result was that one could represent, for example, the combination of this ad hoc category with the concept of Latin America to posit a scenario of an arid, highly developed version of Latin America.

6 Conclusions and Future Work

In this paper we presented a metric conceptual space algebra that defines conceptual spaces as multi-leveled representational structures with operations that are well-suited for computational implementation. A clearer distinction is made between domains and quality dimensions than in previous formalizations of conceptual spaces, which allows for more expressivity for similarity and complex concept combination operations. The conceptual regions that exist in domains are defined as convex polytopes, so that primitive geometric and topological operations can be implemented using algorithms that are tractable. We demonstrated a practical application of conceptual space algebraic operations for a spatial information system, though the algebra presented in this paper is designed for general use [11].

This paper presented a theoretical framework for a metric conceptual space algebra, and there exist many avenues for extending it. The described operations assume no correlation between the different regions that compose a concept, but in reality the quality values of concepts are very often correlated. One possible solution is to integrate multivariate analysis with existing operations. In addition, concepts are often uncertain and dynamic. The algebra described in this paper should be extended to accommodate convex regions that are fuzzy. Parameterized rough fuzzy sets may be one component to such an extension [1]. Introducing operations that allow one to query about the shape of conceptual regions over time would also be a valuable addition to the algebra [26].

Acknowledgments

This work was supported by NSF IGERT Grant #DGE-0221713. The suggestions from four anonymous reviewers helped improve the content of the paper.

References

1. Ahlqvist, O.: A Parameterized Representation of Uncertain Conceptual Spaces. Transactions in GIS, 8(4): 492–514 (2004)
2. Ahlqvist, O., Ban, H.: Categorical Measurement Semantics: A New Second Space for Geography. Geography Compass, 1(3): 536–555 (2007)
3. Aisbett, J., Gibbon, G.: A general formulation of conceptual spaces as a meso level representation. Artificial Intelligence, 133: 189–232 (2001)
4. Barsalou, L.: Situated simulation in the human conceptual system. Language and Cognitive Processes, 5(6): 513–562 (2003)
5. Central Intelligence Agency - The World Factbook. Available at: https://www.cia.gov/library/publications/the-world-factbook/ (2009)
6. Dantzig, G. B.: Linear Programming and Extensions. Princeton University Press, Princeton, NJ (1963)
7. Devore, J., Peck, R.: Statistics - The Exploration and Analysis of Data, Fourth Edition. Duxbury, Pacific Grove, CA (2001)
8. Fukuda, K.: Frequently Asked Questions in Polyhedral Computation. Available at: http://www.ifor.math.ethz.ch/~fukuda/polyfaq/polyfaq.html (2004)
9. Gärdenfors, P.: Conceptual Spaces: The Geometry of Thought. MIT Press, Cambridge, MA (2000)
10. Gärdenfors, P.: How to Make the Semantic Web More Semantic. In: Varzi, A., Vieu, L. (Eds.) Formal Ontology in Information Systems, Proceedings of the Third International Conference (FOIS 2004). Frontiers in Articial Intelligence and Applications, vol. 114, pp. 153–164. IOS Press, Amsterdam, NL (2004)
11. Gärdenfors, P.: Conceptual Spaces as a Framework for Knowledge Representation. Mind and Matter 2, 9–27 (2004)
12. Gärdenfors, P.: Representing actions and functional properties in conceptual spaces, In T. Ziemke, J. Zlatev, R. Frank, (Eds.), Body, Language and Mind. Mouton de Gruyter, Berlin, 167–195 (2007)
13. Gärdenfors, P., Williams, M.: Reasoning About Categories in Conceptual Spaces. in Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence Morgan Kaufmann, 385–392 (2001)
14. Grünbaum, P.: Convex Polytopes, 2nd edition. Kaibel, V., Klee, V., Ziegler, G. (Eds.) Springer, New York (2003)
15. Harnad, S. Categorical Perception: The Groundwork of Cognition. Cambridge University Press, Cambridge, UK (1987)
16. Janowicz, K., Raubal, M.: Affordance-Based Similarity Measurement for Entity Types. In S. Winter, M. Duckham, L. Kulik, B. Kuipers (Eds.), Spatial Information Theory - 8th International Conference, COSIT 2007, Melbourne, Australia, September 2007 (Vol. 4736, pp. 133–151). Springer, Berlin (2007)
17. Janowicz, K., Raubal, M., Schwering, A., Kuhn, W.: Semantic Similarity Measurement and Geospatial Applications. Transactions in GIS 12(6): 651–659 (2008)
18. Johannesson, M.: The Problem of Combining Integral and Separable Dimensions. Lund University Cognitive Studies, Lund, Sweden, 16 (2002)

19. Keßler, C.: Similarity Measurement in Context. B. Kokinov, D. Richardson, T. Roth-Berghofer, L. Vieu (Eds.) CONTEXT 2007. Springer Lecture Notes in Artificial Intelligence 4635, pp. 277–290 Berlin (2007)
20. Lakoff, G.: Cognitive Semantics, In U. Eco, M. Santambrogio, P. Violi (Eds.), Meaning and Mental Representations. Indiana University Press, Bloomington, 119–154 (1988)
21. Mark, D., Freksa, C., Hirtle, S., Lloyd, R., Tversky, B.: Cognitive models of geographical space. International Journal of Geographical Information Science 13(8): 747–774 (1999)
22. Montello, D., Freundschuh, S.: Cognition of Geographic Information. In R. McMaster, E. Usery (Eds.), A research agenda for geographic information science. CRC Press, Boca Raton, FL 61–91 (2005)
23. Nocedal, J., Wright, S. J.: Numerical Optimization. Springer, New York (1999)
24. Rasch, G.: On General Laws and the Meaning of Measurement in Psychology. In Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, IV. University of California Press, Berkeley 321–334 (1961)
25. Raubal, M.: Formalizing Conceptual Spaces. In Varzi, A., Vieu, L. (Eds.) Formal Ontology in Information Systems. Proceedings of the Third International Conference (FOIS 2004), 153–164. IOS Press, Amsterdam (2004)
26. Raubal, M.: Representing Concepts in Time. In C. Freksa, N. Newcombe, P. Gärdenfors, S. Wölfl (Eds.) Spatial Cognition VI - Learning, Reasoning, and Talking about Space. Proceedings of the International Conference Spatial Cognition 2008, Freiburg, Germany. Springer Lecture Notes in Artificial Intelligence 5248, pp. 328–343 Berlin (2008)
27. Rickard, J. T.: A concept geometry for conceptual spaces. Fuzzy Optimal Decision Making 5: 311–329 (2006)
28. Rosch, E.: Cognitive Representations of Semantic Categories. Journal of Experimental Psychology: General 104: 192–232 (1975)
29. Schwering, A.: Approaches to Semantic Similarity Measurement for Geo-Spatial Data: A Survey. Transactions in GIS 12(1): 5–29 (2008)
30. Schwering, A., Raubal, M.: Measuring Semantic Similarity between Geospatial Conceptual Regions. In A. Rodriguez, I. Cruz, M. Egenhofer, S. Levashkin (Eds.), GeoSpatial Semantics - First International Conference, GeoS 2005, Mexico City, Mexico, November 2005 (Vol. 3799, pp. 90–106). Springer, Berlin. (2005)
31. Shepard, R. N.: Toward a universal law of generalization for psychological science. Science 237: 1317–1323 (1987)
32. Stevens, S.: On the Theory of Scales of Measurement. Science 103: 677–680 (1946)
33. Tanasescu, V.: Spatial Semantics in Difference Spaces. In S. Winter, M. Duckham, L. Kulik, B. Kuipers (Eds.), Spatial Information Theory - 8th International Conference, COSIT 2007, Melbourne, Australia, September 2007 (Vol. 4736, pp. 96–115). Springer, Berlin (2007)
34. Tversky, A.: Features of Similarity. Psychological Review 84(4): 327-352 (1977).
35. UNDP (United Nations Development Program): Human Development Report 2007/2008. Palgrave Macmillan, New York (2007)
36. Winter, S., Raubal, M., Nothegger, C.: Focalizing Measures of Salience for Wayfinding. In L. Meng, A. Zipf, T. Reichenbacher (Eds.), Map-based Mobile Services – Theories, Methods and Implementations (pp. 127-142). Berlin: Springer. (2005)
37. World Resources Institute: EarthTrends: Environmental Information. Available at http://earthtrends.wri.org. World Resources Institute, Washington, D.C. (2006)