Semantic Referencing – Determining Context Weights for Similarity Measurement

Krzysztof Janowicz[1], Benjamin Adams[2], Martin Raubal[3]

[1] jano@psu.edu
Department of Geography, The Pennsylvania State University, USA
[2] badams@cs.ucsb.edu
Department of Computer Science, University of California, Santa Barbara, USA
[3] raubal@geog.ucsb.edu
Department of Geography, University of California, Santa Barbara, USA

Abstract. Semantic similarity measurement is a key methodology in various domains ranging from cognitive science to geographic information retrieval on the Web. Meaningful notions of similarity, however, cannot be determined without taking additional contextual information into account. One way to make similarity measures context-aware is by introducing weights for specific characteristics. Existing approaches to automatically determine such weights are rather limited or require application specific adjustments. In the past, the possibility to tweak similarity theories until they fit a specific use case has been one of the major criticisms for their evaluation. In this work, we propose a novel approach to semi-automatically adapt similarity theories to the user's needs and hence make them context-aware. Our methodology is inspired by the process of georeferencing images in which known control points between the image and geographic space are used to compute a suitable transformation. We propose to semi-automatically calibrate weights to compute inter-instance and inter-concept similarities by allowing the user to adjust pre-computed similarity rankings. These known control similarities are then used to reference other similarity values.

Keywords: Semantic Similarity, Geo-Semantics, Information Retrieval

1 Introduction and Motivation

Similarity and analogy based reasoning are major approaches for the understanding of human cognition [1], work on artificial intelligence [2], as well as information retrieval and knowledge organization. In his classic book *Gödel, Escher, Bach - An Eternal Golden Braid*, Hofstadter, for instance, lists among the fundamental building blocks of human intelligence the ability *to find similarities between situations despite differences which may separate them [and] to draw distinctions between situations despite similarities which may link them* [3, p. 26]. The power of similarity lies in providing a graded structure instead of a rigid Boolean matching. In contrast to many purely syntactical or statistical

measures, semantic similarity computes proximity based on the meaning of compared terms. Semantic similarity measures have a long tradition in GIScience – partially due to the analogy between measuring distances in geographic space and computing semantic similarity as inverse distance within a semantic or conceptual space [4]. Over the years, these measures have been applied to compute the similarity between spatial scenes [5,6], to improve landmark-based navigation [7], browsing through digital gazetteers [8], to support the classification of remote sensing data [9], or as additional reasoning service to compare or align classes, instances, and terms on the (geospatial) semantic Web [10,11,12,13,14]. The meaning of terms, however, is influenced or even determined by the context in which they are uttered. Therefore, meaningful similarities cannot be determined without taking additional contextual information into account [15,16,17]. A classical approach to make similarity theories context-aware is by introducing flexible weights. Most existing approaches to determine these weights are either too broad, application specific, or do not take users and their requirements into account [16]. In the past, the ability to adjust similarity theories until they fit a specific purpose has been one of the major criticisms for their evaluation.

In this work, we introduce a novel, context-aware, and semi-automatic weighting approach to better approximate the user's needs. The proposed methodology is inspired by the process of georeferencing images in which known control points between an image and the geographic space determine the appropriate transformation. In analogy, we propose to calibrate weights by allowing users to adjust the similarity of prominent pairs in a ranking. Some of these adjustments can be done automatically, for instance, by taking the user's location into account. Based on the user's *control similarities*, we can adjust the weights and hence reference other similarity values within an ontology. In analogy to geo-referencing, we call this process *semantic referencing*. Note, however, that in fact our approach is an optimization task.

The remainder of this paper is structured as follows. First we introduce related work on semantic similarity and information retrieval. Then, we discuss the relation between feature-based and geometric theories from on ontological perspective. Next, we introduce the theory of semantic referencing and diagnosticity measures. We then demonstrate our approach and its limits using an example from forestry and conclude the paper by pointing to open research questions.

2 Related Work

This section introduces related work on semantic similarity measurement and provides a definition for information retrieval.

2.1 Semantic Similarity

Similarity, the degree to which entities, concepts, or scenes resemble one another, is a foundational topic in many areas of cognitive science [1]. Semantic similarity measurement refers to the process of calculating an interval scaled value of

the proximity of the meanings. The importance of similarity measurement for categorization has been demonstrated through the observation of prototype effects, which show that objects are classified based on their semantic distance to an idealized prototype [18]. In GIScience semantic similarity has been an increasingly important topic, especially with respect to geographic information retrieval and the geospatial semantic Web [10]. Context awareness is an important (though often overlooked) component to any cognitively plausible similarity theory [11,16,17]. One approach to identify saliency weights for a given context was introduced by Tversky [19]. This approach uses the notion of *diagnosticity*, which indicates that the entity set being compared has a diagnostic effect of making certain features more salient with respect to their similarity. Rodriguez and Egenhofer [11] introduced the Matching-Distance Similarity Measure (MDSM) for measuring the similarity of geospatial features represented in a feature-based ontology. MDSM incorporates diagnosticity by utilizing the variability and commonalities of features in the ontology to determine their salience weights. These weights are then used to compute similarity. The Sim-DL similarity server implements a context-sensitive measure for concepts specified in description logics used on the semantic Web [12]. SIM-DL automatically adjusts the similarity of relations and primitive concepts based on context parameters provided by the user. Raubal [7] formalizes context in a similarity measure for geometric conceptual space representations by applying weights to the individual dimensions.

2.2 Geographic Information Retrieval

While information retrieval is an interdisciplinary research field including work on indexing and data storage, we focus on the relevancy relationship used to judge whether discovered information matches the user's needs. Formally, as shown in equation 1, information retrieval is about the degree of relevance between an object or set of objects and information desired by the user. The information sought is specified not only by an explicit query but also by implicit and inferred information gleaned from the user (e.g., a formal representation of personalization variables) [20]. In the context of Geographic Information Retrieval (GIR) the implicit information is spatially context-sensitive [21]. For example, the location of the user gives a GIR application implicit information that can be used to refine or otherwise alter search results to a local area.

$$IR = m[\mathfrak{R}(O, (Q, \langle I, \mapsto \rangle))] \tag{1}$$

where

- $\mathfrak{R}$ is the relevance relationship,
- O is a set of objects,
- Q is the user's query,
- I is implicit information,
- $\mapsto$ is inferred information,
- and m is the degree (or certainty) of relevance.

For the purpose of this work we use similarity as relevance relationship, while the query and compared-to objects are concepts from geo-ontologies.

3 Reification and Similarity

There are at least five major approaches to semantic similarity measurement, those based on computing feature overlap, on counting transformation steps, finding alignments, computing graph-distance in a network, and those based on geometric spaces; see [1,22] for recent overviews. As each of these approaches has its benefits and drawbacks, most modern similarity theories combine them to increase expressivity. For instance, similarity theories based on features or geometry are limited in their ability to handle relationships and are therefore enriched by network-based measures to form a hybrid model [11,23,12]. While there has been some work on translating and combining feature-based and geometric approaches in cognitive science [24,25], this topic has not received much attention in ontology engineering so far – a notable exception being work on ontological design patterns [26]. A classic translation example is the representation of dimensions such as length by sets of nested features [25].

From an ontological perspective, and as recently proposed by Scheider et al. [27], we argue that features are *fictions* which result from reifications[1] of (directly) perceivable observations. This shifts the debate from computational aspects to questions of granularity. Just like with the representation of geographic features, such as cities or transportation infrastructures, changes happen from points and polylines to polygons with scale, the description of concepts changes from features to regions in geometric space. This is the same process as applied in creating feature hierarchies or complex dimensions[2]. An ontology of land-use may list `Afforested` as feature type, while on a more detailed level the same notion can be modeled as minimum percentage value on the `CrownCover` dimension.
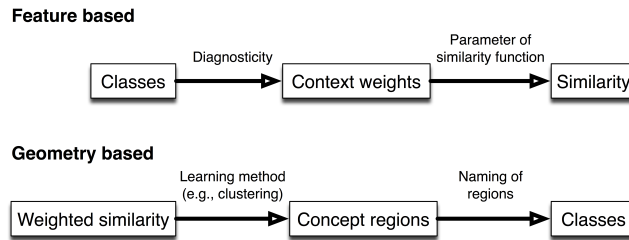
From the perspective of similarity measurement, we can regard feature-based similarities as coarser grained versions of similarities computed by geometric approaches. Hence, we can switch between them depending on the required granularity (as long as we can re-reificate the features). Note, however, that as feature-based similarity computes overlap while geometric approaches compute distance in a vector space the semantics of similarity changes with the translations.

The role context plays with respect to the relation between similarity and classification differs depending on whether a feature-based or geometric representation is used (figure 1). Diagnosticity in the feature-based representation assumes an *a priori* classification to determine weights on features that in turn are used as inputs for measuring similarity. Hence, classification generates similarity. The geometric approach, in contrast, is the opposite – classification is the result of a distance-based similarity function where context is represented by saliency weights on the dimensions. Though the dimensions used to measure similarity may be chosen *a priori*, the regions that represent the classes are not (see also [28]). Thus, the definition of classes depends on the context. These dif-

---

[1] We restrict the notion of reification to the objectification of relations or dimensions.

[2] This is similar to the shift from prototypes modeled as points in a geometric space towards regions. However, this involves slightly changing semantics.

Feature based

Classes —Diagnosticity→ Context weights —Parameter of similarity function→ Similarity

Geometry based

Weighted similarity —Learning method (e.g., clustering)→ Concept regions —Naming of regions→ Classes

**Fig. 1.** Relationship of similarity and classification in different representations.

ferences have important implications for switching between granularities. If we consider features as analogous to regions in the geometric representation then reification can be thought of as the *labeling* of a particular classification (for a particular context) in the geometric representation.

4 Semantic Referencing

Semantic similarity measures are especially beneficial for navigating and browsing through large knowledge bases, i.e., for information retrieval, as well as for ontology engineering. They can be used to reduce the burden of understanding formal definitions [29], are more flexible than rigid (keyword) matching approaches, and help establishing new relationships between information. All these use cases, however, require that the similarity measures are cognitively plausible, i.e., that their rankings correlate with those from human users. As similarity is highly context sensitive, most recent similarity theories implement various context models [16]. While (semi) automatic weights are applied in many cases, they can only roughly approximate the similarity drift caused by additional information which is not explicitly stated in the user's query; see section 2.2.

A promising approach would be to combine weights with additional user feedback, i.e., allow users to influence the weighting process. However, assigning weights for features or dimensions of large ontologies would be a time consuming and error-prone task. To have perfect information about the user's preferences would require manual weights for all features and dimensions; taking asymmetry into account would even double the number of required pairwise comparisons. Finally, a user would have to take abstract decisions such as weighting the similarity between `Afforested` and `Artificial` which both may be features in a forestry ontology. Consequently, a feasible solution has to infer weights from partial information. In principle, there is an infinite number of possible context weights and their combination, which shifts the problem to an optimization challenge related to classical work from multi-criteria analysis.

In previous work, we have shown how users and domain experts can compare their own similarity estimations to the rankings produced by a similarity server to estimate whether the investigated ontology fits their purpose [29]; see figure
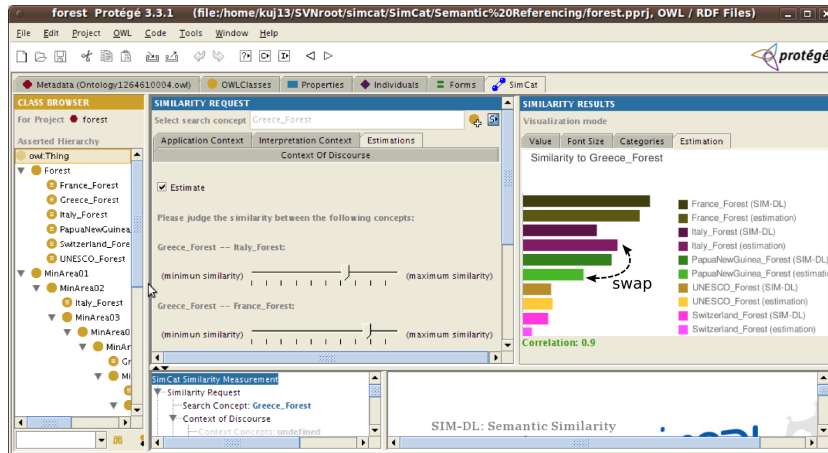
Fig. 2. Comparing the SIM-DL similarity estimations with those made by the user. Users should be able to swap ranks and hence influence weights.

2. So far, this approach had two shortcomings that could not be resolved. First, the estimations were done by computing rank-correlations (or concordance and rank-correlation in case multiple users were involved) which are not necessarily cognitively plausible. For example, the relative position in the ranking was not taken into account. Second, the system could only tell the domain experts whether the ontology potentially reflects their views or not, but did not offer a way to adjust the similarity weights produced by the similarity reasoner.

In this work, we propose a method to overcome both shortcomings. First by replacing Spearman's rank correlation coefficient with the *DIR* measure, and second by allowing the users to swap ranking positions to adjust weights semi-automatically; see *swap* in figure 2. DIR is a cognitively plausible dissimilarity measure for information retrieval result sets. It is based exclusively on result rankings and therefore applicable independent of the retrieval method. Unlike statistical correlation measures, DIR reflects how users quantify the changes in information retrieval result rankings [17]. It is defined as a symmetric function, which calculates the shift every concept undergoes when a query is posed in different contexts. A weighting function insures that shifts at the top of the rankings are emphasized. Note that we do not present abstract features or dimensions to the user but selected concepts ranked by their similarity. Moreover, the user does not need to take pair-wise decisions but directly changes the position of target concepts in the similarity ranking. These changes are then used to adjust the feature or dimension weights.

In analogy to georeferencing we call the process in which the weights get re-computed based on partial information provided by the user the *semantic referencing of similarities*. Georeferencing is the act of identifying a direct or indirect relation between an entity and a geographical position in space. In photogram-

metry, control points on the ground are used to fix the scale of the photographs. This can be simply done by measuring the natural distance between two points on the ground that can also be identified on the photographs. If a high degree of accuracy is required, then premarked points on the ground rather than natural features are used and based on the ground and picture coordinates a transformation is calculated [30]. While it is useful to think of the user adjustments as a kind of known control similarities (adequate to the user's conceptualization), there are also clear differences between both methodologies. These are grounded in the fact that semantic referencing has to cope with an arbitrary number of dimensions, not all of them can be adjusted by swaps in a single ranking, and that human notions of distance do not necessarily fulfill the metric requirements.

In the following we introduce a basic algorithm schema and diagnosticity functions for feature-based approaches. To demonstrate that our approach is generalizable we discuss which extensions are necessary for geometric models. We also introduce variability and commonality measures for geometric similarity measures which have not been investigated so far.

4.1 Semantic Referencing for Feature-based Similarity Measures

For reasons of simplification, and in accordance with the classical feature-based theories, we assume concepts are defined by the intersection of more primitive ones which in turn can be further decomposed into features. Consequently, we leave logical negation, disjunction, and relationships between concepts and individuals aside. In such a representation language role-filler pairs can be represented as single features such as `NextToTransportationInfrastructure` which, as argued above, are reifications. For lack of space, we reduce the feature-based similarity between concepts to a ratio of common versus distinct features leaving asymmetry aside. Moreover, we only discuss *commonality* as diagnosticity measure and leave *variability* aside. The notion of asymmetry in Tversky's contrast model and variability have been extensively discussed in the literature [11]. Both can be included in the presented algorithm without major modifications, e.g., variability is just the inverse of commonality.

We assume that a user defines a query Q, in our case by selecting a search concept C_s, using a graphical user interface; for instance the SIM-DL Protégé plug-in or the semantics-based gazetteer web interface [16,8]. Instead of setting fixed weights, the task is to infer ($\mapsto$; see section 2.2) the weights from (explicit and implicit) information (I) provided by the user. We further assume that O, the set of objects in the information retrieval definition, is a set of target concepts $c_{t1}, ..., c_{tn}$ from the examined ontology. Different solutions have been proposed to determine which concepts should serve as search and target concepts. The Literature about multi-criteria analysis with partial information proposes to take examples in which the users are experts. Others propose to use the concepts which have been mentioned and grouped together most often during ontology engineering and knowledge acquisition tasks [29]. For populated ontologies, those concepts with the highest count of individuals may be a good choice as they have the highest probability to be used subsequently.

Listing 1.1. Basic Algorithm; one swap per turn version.

```
1   for (feature : ⋃C_{t_i} ∪ C_s)
2     computeGlobalDiagnosticity(                    // See equations 3, 10.
3       computeLocalDiagnosticity(feature)); // See equations 2, 6, 8, 9.
4
5   for (C_t : O)
6     sortByDescendingDiagnosticity(C_t);
7     computeSimilarity(C_s, C_t);
8
9   theoryRanking = sortByDescendingSimilarity(O);        ⟶ [USER]
10  humanRanking = retrieveRanking();                     ⟵ [USER]
11
12  if (computeDIR(theoryRanking, humanRanking) ≤ threshold)
13    terminate; // No weight adjustment required.
14
15  candidateFeatureList(O); // See equation 4 and section 4.2.
16  modifyDiagnosticities(
17    humanRanking.swapFrom, humanRanking.swapTo); // See equations 5, 11.
```

Listing 1.1 shows the main steps to readjust the weights according to the ranking changes proposed by the user. First, the diagnosticity of each feature (f) has to be computed and normalized: see equations 2 and 3.

$$localDiagnosticity(f) = \frac{|\{f | f \in \bigcup C_{t_i} \cup C_s\}|}{\{|\bigcup C_{t_i} \cup C_s|\}} \tag{2}$$

$$globalDiagnosticity(localDiagnosticity) = \frac{localDiagnosticity}{\sum_{f_1}^{f_n} localDiagnosticity(f_i)} \tag{3}$$

Next, the features for each concept definition are ordered by their diagnosticity and the similarity for all search concept - target concept pairs is computed using a feature-based (or geometric) theory. As shown in lines 8 and 9, the resulting similarity ranking is presented to the user who can decide to swap two positions in the ranking (per turn). The user interface has to give the user the possibility to actively move a concept up or down as the choice of directions matters, i.e., the changes are asymmetric. Next, the DIR measure is used to determine how dissimilar both rankings are and whether an adjustment is necessary.

$$candidateFeatureList(O) = \{f | (f \in C_s) \wedge ((f \in C_{sf}) \bigoplus (f \in C_{st}))\} \tag{4}$$

As indicated in equation 4, only those features are candidates for weight adjustment which appear in the source and one of the target concepts (but not both). This is not required for theories which support similarity *between* features/dimensions such as SIM-DL or geometric approaches which will be discussed below.

$$
\begin{aligned}
MOD : \ & (x \pm C_{sf}[0].globalDiagnosticity) * sim(C_s[y], C_{sf}[0]) \\
& + (\frac{x}{2^1} \pm C_{sf}[1].globalDiagnosticity) * sim(C_s[y], C_{sf}[1]) \\
& + (\frac{x}{2^2} \pm C_{sf}[2].globalDiagnosticity) * sim(C_s[y], C_{sf}[2]) + ... \frac{x}{2^n} \\
& = sim(C_s, C_{st}) + 0.01;
\end{aligned}
\tag{5}
$$

Equation 5 shows how the weights are increased (+) or decreased (-) based on the user's modifications where C_{sf} is the swapped-from and C_{st} the swapped-to concept in the ranking. y is the to-be-compared feature in the C_s list. Instead of arbitrary changes to multiple weights we use a power function to model the Max Effect described in cognitive science studies [31]. This effect describes the tendency to favor a particular reasoning strategy which has turned out to be successful in previous similarity estimations and could be compared to the Matthew Effect in social science. Features which are more diagnostic gain even more diagnosticity while the diagnosticity of others increases slower. So far our algorithm considers local optimizations. It tries to ensure that the weights reflect the swap in the first place and put less emphasis on other parts of the ranking (which can still be adjusted in the next turn). Note that modifying the weights does not always successfully change the ranking or guarantees that the process converges at all. This is especially the case if the user's initial conceptualization differs clearly from the computational representation in the ontology[3] or if users take irrational decisions. In this case the ontology is unsuitable for the given context. After the adjustment, the new diagnosticity weights are used within the ontology to better approximate the user's preferences.

4.2 Measuring Diagnosticity in Geometric Representations

In this section we extend the MDSM measures of commonality and variability to geometry-based representations.

Case 1 - Commonality We have a set of concepts that are represented as incomplete vectors (i.e., points) in a continuous multidimensional space. That is, for any given concept values may be undefined for one or more of the dimensions. In this case we are interested in determining the diagnosticity of each dimension based on how many concepts have a value defined for it. Let $C = \{c_1, c_2, \ldots, c_n\}$ be the set of concepts. The probability $p(c, d)$ that a concept $c \in C$ has a value defined for a given dimension d is equal to the number of concepts defined for d over the total number of concepts in C. The *commonality* diagnosticity (CD) of a dimension d is thus defined in equation 6.

$$CD = \frac{p(c, d)}{\sum_{i=1}^{m} p(c, d_i)} \tag{6}$$

Case 2 - Variability We have a set of concepts that are represented as complete vectors in a continuous multidimensional space. In this case, the concepts are defined using the same dimensions, but they still differ semantically in that they are represented by different points in the space. This situation will occur

---

[3] The user does not need to know or understand the formal definitions of concepts displayed in the ranking which at the same time is a major benefit of our approach as discussed before.

when representing a set of observations construed as exemplars of a concept (or in ontological language an enumeration of individuals) and that are measured using the same methods.

Since there are no differences between the dimensions, diagnosticity is measured in terms of deviation of the data along the different dimensions. The semantic interpretation of the deviation depends on whether the data points represent different exemplars of the same concept (e.g., different definitions of forest) or in fact different concepts (e.g., forest, woodland, chaparral, etc.). Our hypothesis is that if a set of instances of the same concept varies little along one dimension (x) and a lot along another dimension (y) then dimension x is more salient and therefore more diagnostic. In the case of different concepts the opposite is true. Intuitively, we want to identify which quality values are most alike for the instances of the same concept and which ones help us to distinguish between different concepts.

The proposed method is to compare the mean absolute deviation (MAD) of the data values (X) for each dimension. To maintain consistency with MDSM terminology we call the MAD value of a dimension its *variability*.

$$MAD = \frac{1}{n} \sum_{i=1}^{n} |x_i - m(X)| \tag{7}$$

The data must first be normalized to [0,1] along each dimension so that the MADs can be compared. Depending on the data, different normalization techniques may be necessary. In general, a Min-max normalization will be sufficient, though in the case that the data have a fixed range then the range minimum and maximum is preferred (e.g., any ratio scaled dimension will have a minimum value of zero). It is noted, however, that Min-max normalization is sensitive to outliers. Equation 8 shows the *variability* diagnosticity (VD) of a dimension for the case when different instances of the same concept are represented.

$$VD_{same} = \frac{1 - \frac{MAD(d)}{\sum_{i=1}^{m} MAD(d_i)}}{m - 1} \tag{8}$$

VD in the case of different concepts is defined in equation 9.

$$VD_{diff} = \frac{MAD(d)}{\sum_{i=1}^{m} MAD(d_i)} \tag{9}$$

Combining Commonality and Variability The scenario presented in case 1 will always include different distributions of values along each of the dimensions as well, so we define a diagnosticity measure for a given dimension d (equation 10) that combines the two measures listed above. However, the VD measure (see equations 8 and 9) is changed slightly to ignore any undefined values when calculating the MAD.

$$diagnosticity(d) = \frac{CD(d) \times VD(d)}{\sum_{i=1}^{m} CD(d_i) \times VD(d_i)} \tag{10}$$

Modifying Diagnosticities for Geometric Representation Equation 11 shows an extended MOD function for calculating new diagnosticities for dimensions in geometry based representations.

$$
\begin{aligned}
MOD \: : \: & (x + diag(dimP[0])) * sim(C_s[dimP[0]], C_{sf}[dimP[0]]) \\
& + (\frac{x}{2^1} + diag(dimP[1])) * sim(C_s[dimP[1]], C_{sf}[dimP[1]] + ... \\
& + (x - diag(dimM[0])) * sim(C_s[dimM[0]], C_{sf}[dimM[0]]) \\
& + (\frac{x}{2^1} - diag(dimM[1])) * sim(C_s[dimM[1]], C_{sf}[dimM[1]] + ... \\
& = sim(C_s, C_{st}) + 0.01;
\end{aligned}
\tag{11}
$$

Contrary to the feature method (see equation 4), the candidate dimension list includes all dimensions D shared by C_s, C_{sf}, and C_{st}, because different concepts may share the same dimensions but vary in terms of the values along those dimensions. The diagnosticities of each dimension can be calculated by any of the methods described above, depending on the application. It compares the similarities of C_{sf} and C_{st} to C_s, and constructs two sorted queues, $dimP$ and $dimM$ based on which of the two target concepts is most similar to the search concept for each given dimension (see listing 1.2). The queues are sorted by how large the difference is between the similarities. These queues are used to generate positive and negative power functions, which are combined and solved to generate new diagnosticities for each of the dimensions.

Listing 1.2. Determining which dimensions increase/decrease diagnosticity

```
1  for (d : D) // for each dimension
2    fromToSimDiff = sim(C_s[d], C_sf[d]) - sim(C_s[d], C_st[d])
3    if (fromToSimDiff > 0) // similarity of C_s to C_sf < C_s to C_st
4      dimP << d, fromToSimDiff  // add to + queue sorted by difference
5    else
6      dimM << d, fromToSimDiff // add to - queue sorted by difference
```

5 Application

Data about forest cover in a country are dependent in part on the definition (i.e., semantics) of *forest* used by that country. In order to compare these data across different countries it is helpful to identify which definitions are more similar to one another. This similarity information can then be used to evaluate the degree to which the forest cover data from different countries are comparable. In this section, we present the usage of the semantic referencing algorithm to semi-automatically calculate the similarity of different *forest* definitions. For this example we use a geometry-based representation, where each forest definition is represented as a point in a three dimensional space. The three dimensions are minimum area, minimum crown height, and minimum tree height[4].

---

[4] Source: http://www.affrc.go.jp/satellite/shokusei/EOSD/Background/Gyde_Lund_Definitions_of_Forest_RAD/DEFpaper.html

5.1 Calculating Diagnosticity

The user first selects control points, which are used to calculate the diagnosticity of each of the dimensions. For this particular example we assume some domain expertise on the part of the user regarding forest definitions, so that they have an internal conceptualization with which they can compare the rankings. Table 1 shows a sample selection of control points along with commonality (CD), variability (VD_{same}), and the combined *diagnosticity*(d) calculations based on 6 sample control points. The minimum crown cover dimension is the most diagnostic as it shows a low variability (which is a high diagnostic indicator when comparing exemplars of the same concept) and high commonality as it is defined for all sample countries. The abstract semantic referencing algorithm can use any of the three diagnosticity measures to identify which dimension is most salient; in this case we use the combined measure.

Table 1. Sample forest definition control points and diagnosticities of dimensions

	Min area (ha)	Min crown cover (%)	Min tree height (m)
France	2.0	10.0	
Greece	0.5	10.0	
Italy	0.2	20.0	
Papua New Guinea	100.0	10.0	5.0
Switzerland		20.0	
UNESCO		40.0	5.0
Commonality (CD)	0.333	0.5	0.167
Variability (VD_{same})	0.092	0.408	0.5
Diagnosticity	**0.096**	**0.642**	**0.262**

5.2 Iterating Through the Semantic Referencing Algorithm

For this geometry-based representation we assume a semantic distance (dissimilarity) between two forest definitions is equal to the weighted Manhattan distance between the two vector representations where the weights are the diagnosticities. Similarity is simply defined as 1 - *semantic distance*. If a dimension is defined for the source forest but not in the target then the distance is considered maximal (i.e., 1 on a dimension normalized to [0..1]). If the dimension is undefined for the source the distance along that dimension is considered to equal zero. Using this measure we calculate a ranking of the target forest definitions to the source target, which in our example is the definition of Greek forest with the following similarity ranking: 1. France; 2. Italy; 3. Papua New Guinea; 4. Switzerland; 5. UNESCO. The user can choose to accept the ranking or adjust it by moving a target up or down. Say the user wants to move Papua New Guinea to Italy's ranking. Using the geometry-based MOD function (see equation 11) the diagnosticity of minimum area is reduced and the diagnosticity of minimum crown cover is increased, because Papua New Guinea's and France's forest definitions are more similar along the minimum crown cover dimension than Italy's and

France's are (and vice versa for minimum area). Similarity rankings are recalculated using the new diagnosticity values and the above process is reiterated until the user gets an acceptable ranking, which is then used for similarity measures on a wider set of forest definitions in a traditional information retrieval setting.

6   Conclusions and Future Work

In this paper we have discussed the relationship between feature-based and geometric similarity theories from the viewpoint of ontology engineering, defined diagnosticity for geometric similarity measures, and provided a novel methodology for adjusting weights based on user preferences. While the user provides an explicit query, the main innovation in the semantic referencing approach is to use implicit information. Instead of presenting abstract feature or dimension pairs, we propose to let the users adjust precomputed inter-concept similarity rankings to learn about their contextual preferences and apply the extracted weights to the ontology.

While this is the first step, our long-term vision is to apply the semantic referencing methodology automatically based on the user's similarity estimations as depicted in figure 2. Such approach would be more intuitive and could be directly integrated into our similarity servers. This, however, would require more sophisticated and global optimization functions. The challenge in developing such functions is not to find *a* working algorithm, but to ensure its cognitive plausibility. It turns out that weighting approaches have their limitations and may not be able to model the user's needs in all cases. The presented work is based on established and well tested measures, complex optimization approaches may, however, require changes in the alignment process or flexible distance metrics differing among dimensions, and hence, will also require extensive human participants test. Moreover, the user's context can also contain additional implicit information to further refining the results. For example, certain characteristics of geographic feature types may be more or less salient depending on the country (and language) of origin of the search. The visualization and interaction with conceptual spaces is also an important field for further research – we believe that parallel coordinate plots may be an interesting solution to some of these challenges.

Finally, there are several improvements to make the diagnosticity measures more robust. First, within the geometric representation there is an assumption that the dimensions are orthogonal allowing us to measure the variability of each dimension independently, but this is not necessarily true if there are correlations between these dimensions. Second, the measure of a domain's variability based on a set of property regions in the geometric representation is simplified to a point measure, which does not consider the structure and size of the regions. One parameter to consider is the degree and kind of overlap among different regions. Other measures of diagnosticity such as the information entropy of a set of data values should be explored and evaluated as well. A region connection calculus can be used to reify the topology of the regions to feature-based properties and

their relations, thus suggesting a strategy for developing a diagnosticity measure that encompasses hybrid feature-geometry representations [32].

Acknowledgments

References

1. Goldstone, R.L., Son, J.: Similarity. In Holyoak, K., Morrison, R., eds.: Cambridge Handbook of Thinking and Reasoning. Cambridge University Press, Cambridge (2005) 13–36
2. Rissland, E.L.: Ai and similarity. IEEE Intelligent Systems **21**(3) (2006) 39–49
3. Hofstadter, D.R.: Gödel, Escher, Bach: An Eternal Golden Braid. Basic Books (1999)
4. Gärdenfors, P.: Conceptual Spaces - The Geometry of Thought. Bradford Books, MIT Press, Cambridge, MA (2000)
5. Nedas, K., Egenhofer, M.: Spatial similarity queries with logical operators. In Hadzilacos, T., Manolopoulos, Y., Roddick, J., Theodoridis, Y., eds.: SSTD '03 - Eighth International Symposium on Spatial and Temporal Databases. Volume 2750 of Lecture Notes in Computer Science., Santorini, Greece (2003) 430–448
6. Li, B., Fonseca, F.: Tdd - a comprehensive model for qualitative spatial similarity assessment. Spatial Cognition and Computation **6**(1) (2006) 31–62
7. Raubal, M.: Formalizing conceptual spaces. In Varzi, A., Vieu, L., eds.: Formal Ontology in Information Systems, Proceedings of the Third International Conference (FOIS 2004). Volume 114 of Frontiers in Artificial Intelligence and Applications. IOS Press, Torino, Italy (November 2004) 153–164
8. Janowicz, K., Schwarz, M., Wilkes, M.: Implementation and evaluation of a semantics-based user interface for web gazetteers. In: Workshop on Visual Interfaces to the Social and the Semantic Web (VISSW2009). (2009)
9. Ahlqvist, O.: Extending post classification change detection using semantic similarity metrics to overcome class heterogeneity: a study of 1992 and 2001 national land cover database changes. Remote Sensing of Environment **112(3)** (2008) 1226–1241
10. Egenhofer, M.: Toward the semantic geospatial web. In: GIS '02: Proceedings of the 10th ACM international symposium on Advances in geographic information systems, New York, NY, USA, ACM (2002) 1–4
11. Rodríguez, A., Egenhofer, M.: Comparing geospatial entity classes: an asymmetric and context-dependent similarity measure. International Journal of Geographical Information Science **18**(3) (2004) 229–256
12. Janowicz, K., Wilkes, M.: $SIM-DL_A$: A Novel Semantic Similarity Measure for Description Logics Reducing Inter-concept to Inter-instance Similarity. In Aroyo, L., Traverso, P., Ciravegna, F., Cimiano, P., Heath, T., Hyvoenen, E., Mizoguchi, R., Oren, E., Sabou, M., Simperl, E., eds.: 6th Annual European Semantic Web Conference (ESWC2009). Volume 5554 of LNCS., Springer (2009) 353–367
13. Cruz, I., Sunna, W.: Structural alignment methods with applications to geospatial ontologies. Transactions in GIS **12**(6) (2008) 683–711

14. Adams, B., Raubal, M.: A metric conceptual space algebra. In Hornsby, K.S., Claramunt, C., Denis, M., Ligozat, G., eds.: Conference on Spatial Information Theory (COSIT 2009). Volume 5756 of Lecture Notes in Computer Science., Springer (2009) 51–68

15. Goldstone, R.L., Medin, D.L., Halberstadt, J.: Similarity in context. Memory & Cognition **25** (1997) 237–255

16. Janowicz, K.: Kinds of contexts and their impact on semantic similarity measurement. In: 5th IEEE Workshop on Context Modeling and Reasoning (CoMoRea'08) at the 6th IEEE International Conference on Pervasive Computing and Communication (PerCom'08). (2008) 441–446

17. Keßler, C.: What's the difference? - a cognitive dissimilarity measure for information retrieval result sets. Knowledge and Information Systems (forthcoming)

18. Goldstone, R.L.: The role of similarity in categorization: providing a groundwork. Cognition **52**(2) (1994) 125 – 157

19. Tversky, A.: Features of similarity. Psychological Review **84**(4) (1977) 327–352

20. Dominich, S.: The Modern Algebra of Information Retrieval. Springer (2008)

21. Keßler, C., Raubal, M., Wosniok, C.: Semantic rules for context-aware geographical information retrieval. In P. Barnaghi, K. Moessner, M.P..S.M., ed.: Smart Sensing and Context Conference 2009. Volume 5741 of LNCS., Springer (2009) 77–92

22. Schwering, A.: Approaches to semantic similarity measurement for geo-spatial data - a survey. Transactions in GIS **12**(1) (2008) 5–29

23. Schwering, A., Raubal, M.: Spatial relations for semantic similarity measurement. In: Perspectives in Conceptual Modeling: ER 2005 Workshops CAOIS, BP-UML, CoMoGIS, eCOMO, and QoIS. Volume 3770 of Lecture Notes in Computer Science. Springer (October 2005) 259–269

24. Navarro, D., Lee, M.: Combining dimensions and features in similarity-based representations. In S. Becker, S.T., Obermayer, K., eds.: dvances in Neural Information Processing Systems 15. MIT Press (2003) 59–66

25. Gati, I., Tversky, A.: Representations of qualitative and quantitative dimensions. Journal of Experimental Psychology: Human Perception and Performance **8(2)** (1982) 325–340

26. Gangemi, A.: Ontology design patterns for semantic web content. In: Proceedings of the Fourth International Semantic Web Conference (ISWC 2005). Volume 3729 of Lecture Notes in Computer Science., Springer (2005) 262–276

27. Scheider, S., Probst, F., Janowicz, K.: Constructing bodies and their qualities from observations. In: 6th International Conference on Formal Ontology in Information Systems (FOIS 2010 forthcoming)

28. Ahlqvist, O.: In search for classification that support the dynamics of science ? the fao land cover classification system and proposed modifications. Environment and Planning B: Planning and Design **35**(1) (2008) 169–186

29. Janowicz, K., Maué, P., Wilkes, M., Braun, M., Schade, S., Dupke, S., Kuhn, W.: Similarity as a quality indicator in ontology engineering. In Eschenbach, C., Grüninger, M., eds.: 5th International Conference on Formal Ontology in Information Systems. Volume 183., IOS Pres (October 2008) 92–105

30. Kraus, K.: Photogrammetry: Geometry from Images and Laser Scans. 2nd edn. Walter de Gruyter (2007)

31. Medin, D., Goldstone, R., Gentner, D.: Respects for similarity. Psychological Review **100**(2) (1993) 254–278

32. Gärdenfors, P., Williams, M.A.: Reasoning about categories in conceptual spaces. In: Proceedings of the Seventeenth International Joint Conference on Artificial Intelligence (IJCAI 2001). (2001) 385–392