Understanding spatial complexities: could similes and metaphors be useful?

A. Holt
Department of Information Science, University of Otago
PO Box 56, Dunedin New Zealand
Email: aholt@infoscience.otago.ac.nz

Abstract. Are there possible benefits to be gained from emulating the human way (of thought and language) with a machine way? This article outlines the concepts of similes and metaphors. This article aims to emphasise that similes and metaphors can be utilised to help understand and extract useful information from spatial data. As similes and metaphors are utilised in human language and understandings it may also be efficacious for spatial domains, in that it may inform and allow us to understand spatial complexities. Similes and metaphors are the verbal expressions of similarity. This article couches similes and metaphors in our spatial language (verbal expressions) as spatial similarity. The author realises that spatial similarity is just one subset of our spatial cognition ability. The general concept outlined in this article is that spatial similarity is a generative technique to analyse the cached information inherent in spatial sciences. This similarity technique is utilised to group natural kinds that manifest similar amenities in complex environments. If we then know or understand one spatial entity in this natural group (s1) then we might understand other entities within that same natural group (s1) from the complex environment.

1. INTRODUCTION

Spatial similarity coalesces researchers working on issues of categorisation (concept representation) in the areas of artificial intelligence, cognitive psychology and geographical information systems. Categorisation[1] may be described as a process by which distinct entities are treated as equivalent. This is one of our fundamental and omnipresent cognitive activities, in that, it allows us to understand and make predictions about objects and events in our world. Our understanding is based on what we know. If "vrystaat", for example, is equivalent to "gjusol" (according to our criteria) then we can predict the behaviour and properties of "vrystaat" (previously unknown) based on the behaviour and properties of "gjusol" (previously known = memory). The problem of understanding what criteria are used to group together entities in the same category is what categorisation is about. Most research has proposed that perceptual or structural similarity is the "glue" that binds objects of a same category. Now a shift has occurred where different and novel propositions have been suggested by psychologists, where maybe objects are grouped together because they facilitate a common goal or serve the same function. Some categories are viewed as coherent because they rest on a theory that explains the commonalities of their elements. The representation of concepts that a categorisation system generates is of course intimately tied to the criteria this system uses to group entities into categories, so along with new models of categorisation, we expect to see the emergence of new models of concept representation apart from the classical ones deriving from the aristotelician, the prototypical and the exemplar views. The representation of the entities to catergorise plays also an important part in the categorisation process. In particular, the context in which the entities occur may influence the way they are classified[2]. The purpose of this paper is to bring fresh insights concerning the notion of spatial similarity, the way goals of categorisation influence this process, how the notion of the theory of a concept can be formalised (using case-based reasoning (CBR)) and implemented in computational models of categorisation and the implications those elements may have on the representation of concepts.

2. SIMILES AND METAPHORS

It is suggested that similes and metaphors are essential ingredients of creative verbal expressions in that they are usually understandable and often informative (Tversky & Krantz 1970; Tversky, 1977). **Similes and metaphors** could also allow us to partially solve the difficulty of dealing with **context** in spatial similarity. That is, when using the same data set, two users can ask for similarities in the data and they could each get different results. Therefore, what is similar for one user may not be similar for another, however they may both have correct results. It should be stressed that the geocomputation techniques is not the similarity result rather, it is the similarity process. Similes and metaphors allow us to get totally different results whilst using the same process, which allows us to deal with different contexts. This article outlines some avenues to use similes and metaphors as a geo-computation tool for comparison and matching of complex entities.

[1] "Categorization and Concept Representation: Models and Implications"http://perso.wanadoo.fr/colette.faucher/categorization.html

[2] "Categorization and Concept Representation: Models and Implications"http://perso.wanadoo.fr/colette.faucher/categorization.html

A simile is "an explicit likening of one thing to another"[3] and "A figure of speech in which two essentially unlike things are compared, the comparison being made explicit typically by the use of the introductory 'like' or 'as' ..." [4]. A metaphor is "a figure of speech by which a thing is spoken of as being that which is only resembles, as when a ferocious man is called a tiger"[5] and a "A figure of speech in which a term is transferred from the object it ordinarily designates to an object it may designate only by implicit comparison or analogy, as in the phrase 'evening of life.'"[6].

2.1 Spatial similes and metaphors

Spatial metaphors as in the user interface concept (Kuhn, 1993) is not what this paper is attempting to emphasis. This paper gravitates to spatial metaphors as a concept for associations/analogies to be made of both knowledge and representations as a data mining/knowledge retrieval geocomputation technique.

Spatial metaphors are not only relevant for human communication but also for objects organisation. People navigate in real life spatial environments based on very vague descriptions and facts. They are also familiar with memorising relationships of objects in spatial terms. Spatial representations are used frequently to convey one or more attributes of the information objects to the user: sorting, grouping and so forth[7].

Spatial metaphors structure many of our complex ideas about life, for example, "Looking back over my life" (recollection), and "looking forward to my vacation" (anticipation). The left/right axis of the body is used in English to express several related ideas: comparison, judgment and equality. It is also important to the center/side contrast.

The Left/Right Contrast[8]

On the one hand	on the other hand.
On one side	on the other side.
From one point of view	from the opposite point of view.

Comparison, choice and judgment between things, as when we say "On the one hand,... but on the other hand..." Because our two hands are both similar and different, this phrase can express either similarity or difference, or both simultaneously. The top/bottom contrast and the front/back contrast, however, indicate more difference than similarity and are thus of a different order. The two sides of an argument is where judgment is relevant, as in deciding between two alternatives. Sometimes we must "come down on one side or the other," but other times we can find a middle ground, a happy medium.

The Front/Back Contrast[9]

Progress	Regress
Looking forward to, prospect	Looking back over, retrospect
In the forefront of my attention	Back of my mind, back burner
Spring forward	Fall back
Avant garde	Derriere garde
Ahead	Behind

Can these concepts we used in a GIS coverage/theme sense?

3. CASE RETRIEVAL

A CBR cycle may be described by the following four processes (Aamodt & Plaza, 1994).

1. retrieve the most similar case(s),
2. reuse the information and knowledge in that case to solve the problem,
3. revise the proposed solution,
4. retain the parts of this experience likely to be useful for future problem solving.

This article focuses on the first process in the CBR cycle. Reminding a user of a previous situation (case), can help solve a new problem. Case retrieval is the implementation of a process for finding the set of most similar cases that match the user's current situation. Similarity assessment is the process in which feature values of the new situation are compared to the feature values of each case in the case-base. The cases that are determined to have a similarity value above a stated threshold are retrieved using a similarity assessment metric.

CBR has been used in the GIS community (Benwell, *et al*. 2000); (Holt, A. 1998; 1999; 2000); (Holt & Benwell, 1996; 1999); (Kaster *et al*. (2000); (Macgillivray *et al*. 1999); (Ohlemüller, *et al*. 2000); (Shi & Yeh, 1999); (Yeh & Shi, 1998; 1999). There are many important areas to research using CBR and GIS, for example, adaptation, indexing, knowledge representation, integration with other learning methods, collaborative agent architectures adaptive interfaces, user modelling and visualisation techniques, cognitive models and psychological evaluations. This articles focuses on the spatial similarity research. Similarity assessment is an important step in CBR system. At present, there are no CBR systems that dealt with topological/GIS relations. *Spatial similarity* can be used to help to retrieve;

- Cases according to their geographic location
- Cases according to their topological relations
- Cases according to their abstract values.

---

[3] Chambers's Twentieth Century Dictionary
[4] American Heritage Dictionary
[5] Chambers's Twentieth Century Dictionary
[6] American Heritage Dictionary
[7] MetaSelf:A Visual Aid to Being Human, 1995, P. Carleton http://www.metaself.org/model/main.html
[8] http://www.metaself.org/model/lericon.html
[9] http://www.metaself.org/model/frbacon.html

4. SIMILARITY ASSESSMENT AND METRICS

A metric is a measure. Similarity is a description of how similar one thing is to another. A similarity metric is a measure of how similar a prior case is to a new situation. Similarity assessment is the act of using a specific metric for retrieval. The similarity metrics used in CBR research are diverse. This following section outlines how similarity assessment can be determined using a CBR package called esteem. A spatial slant will be illustrated. The author has used other CBR packages that provide slightly different similarity metrics.

The developer/user has a wide range of options to create a similarity assessment metric based on the esteem similarity assessment process. So what assessment process should be used? Each application is different, and each application is intended for an audience of varying skills and requirements. The developer should consider the skill-level of the end-user and tailor the type of feature matching and assessment method accordingly. For example if the end-user is likely not to be computer literate then a *partial word match* (case insensitive) would give the broadest coverage at the feature level. *Simple feature counting*, as the overall assessment method would most likely give the broadest treatment to the cases in the case-base and give the end-user a smaller selection of more exact matched cases to choose from. Combinations of the other forms of matching and similarity assessment lead to greater precision in case retrieval, but greater complexity of use.

The following text describes in detail the processes used in each type of similarity assessment used in esteem. With the *similarity definition editor* (see figure 1) the developer defines the method and various metrics for determining similarity during case-base retrieval. The editor has these options:

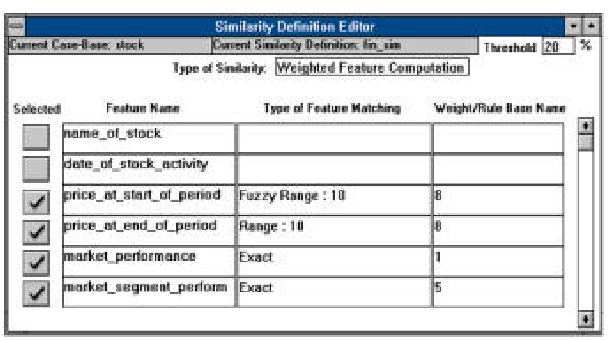

Figure 1: Similarity definition editor

The Similarity Definition editor[10] can be divided into five sections;

4.1 Section one

The buttons on the left of the editor are used to tell the system which features are to be used in similarity matching.

4.2 Section two

The feature names are displayed and then are followed to the right by a field for the specification of the type of feature level matching. A developer can decide how to match at the level of one feature value to another. The developer would right-mouse click within this field to bring up a menu of *Types of Feature Matching*. The type of feature matching available is dependent on what type the feature is.

4.2.1 Exact Feature Value Match

Exact feature match is an exact comparison of: letter by letter, number by number, word by word, list value by list value. Exact feature matching returns a similarity value of 0 or 1.

4.2.2 Partial Value Match

Partial value match is the match of a majority of the feature. Partial value match returns a value between 0 and 1. For example,

Target Feature Value: Geology matches to
Case Feature Value: Gelogy

Comparison of the two values produces a similarity assessment value of 0.86. This is computed by the equation: (1 - number of misspellings) divided by number of characters in the longer word.

4.2.3 Partial Word

Partial word matching compares two values that are strings. The strings are parsed into individual words and a "partial value match" process is applied. The order of the words in each string is disregarded. For example,

Target Value: "Sandy Clay Loam" *matches to*
Case Value: " Sandy Loam Clay"

Produces a similarity value of 1, because order is not considered.

Another example:

Target Value: "Coarse Sandy Loam" matches to:
Case Value: " Sand Clay Loam"

Produces: a similarity value of 0.71, because "Sandy" and "Sand" were off by one character, and the phrases were off by one word "Coarse" and "Clay".

4.2.4 Exact (Case Indifferent)

Exact, case indifferent, provides exact matching regardless of the case of the letters in the values. For example:

Target Value: "Brisbane City" would match to:
Case Value: " brisbane city"

4.2.5 Partial (Case Indifferent)

Partial feature match provides partial feature matching but is case indifferent. For example:

Target Value: "Brisbane City" would match to:
Case Value: "brissy city"

4.2.6 Partial Word (Case Indifferent)

10 User manual ESTEEM Version 1.3

Partial word, case indifferent, would provide matching for partial word values

that have difference in case: For example:

Target Value: "Alec Holt Flew Jumbo Airlines" would match:

Case Value: "alec holt jumbo AirLINES"

4.2.7 Equal Feature Matching

Equal feature matching is used to describe numeric equality. For example:

Target Value: 3 is equal to:

Case Value: 3 as well as:

Case Value: 3.0

4.2.8 Range Feature Matching

Range Feature Matching can be used to describe matches between numbers that have a specified tolerance. The value returned is always 0 or 1. For example: Feature Temperature, Value Definition: Numeric with Range [10 percent] (this implies that values which are within plus or minus 10 percent are considered a match).

Target Feature Value: 40 is a match to:

Case Feature Value: 36.5 and also:

Case Feature Value: 42.3

Similarity value returned is 1

4.2.9 Fuzzy Range Feature Matching

Fuzzy Range Feature matching is used to describe "closeness of match" between numeric values. The similarity score, used during retrieval, changes based on how close or far away the values are from each other. The score changes up to a point. A developer can select the tolerance for fuzzy matching as the percentage of distance one value is from another. Fuzzy Range Feature matching provides a partial match capability for numeric features. Value is a number between, and including 0 and 1. For example: A tolerance of 10 percent is entered

Target Value: 100

Case Value: 97

Values differ by 3 percent. The example uses a tolerance of ten percent, so the returned similarity value is 7 (70 percent).

4.2.10 Inferred Feature Match

Choosing this menu option for the type of matching to perform against a single feature implies that the developer has created a rule-base definition and a set of rules to infer whether one feature value matches another feature value. Using alternative methods of processing the two values would use a set of one or more rules to compare the values. For example:

Feature Value: "sea level" is a match to:

Feature Value: "height=o" if there is a specific rule stating that "sea level " is the equivalent of "height=o ". Or other rules might come into use to compare other feature values to determine whether a target case feature value matches a candidate case feature value.

4.2.11 Recursive Feature Matching

If a feature is of type Case, then the similarity match type of Recursive must be selected. To assess the similarity of two values of that feature, the similarity definition of the cases that represent the values is used. For example, consider a feature called Question1 whose type is case and whose values come from the Questions Case Base. Say two cases, case1 and case2, exist with values for the Question1 feature of QuestionA and QuestionB, respectively. When determining the similarity of case1 and case2, to determine the similarity of the values QuestionA and QuestionB, the current similarity definition of Questions is used.

4.3 Section three

The fields on the right of the editor, next to the scrollbars, are used to either provide a weight that describes the relative importance of that feature in the retrieval process or a rule-base name is given. The rule-base name is a pointer to a collection of rules used to determine how important that feature will be in the retrieval process.

4.4 Section four

The Type of Similarity field is for defining what similarity assessment techniques will be used on a case-by-case basis. Currently ESTEEM supports the following methods of similarity assessment

4.4.1 Feature Counting

This is usually a default form of similarity assessment. Feature counting is a similarity assessment method that is used to compute the closest match between a new situation (target case) and one or more cases from the case-base. A comparison score is computed for all cases by comparing each feature value of the target case with each corresponding feature value for each case in the case-base. The case, or cases, with the highest number of matches is determined to be the most similar. The default weight for each selected feature, to be searched on, is 1.0 and does not play a role in the computation of similarity. The developer can alter the value for weight for any given feature but it does not affect the outcome.

4.4.2 Weighted Feature Computation

A weighted feature computation can be obtained by assigning a value of "importance" to each feature. In general, retrieval of the most relevant case is determined by the presence of a greater number of higher priority (more important) features matching between the target case (new) and the selected case (retrieved).

4.4.3 Inferred Feature Computation

This is another developer-selected form of similarity assessment that uses rules about the domain to determine strength of similarity between a new situation (target case) and the case-base. Inferred feature computation uses rules to compute the weight for a given feature. Based on the values of the new situation's (target case), and specific rules about the domain, the system can compute a value for the weight to be used for matching. Inferred feature matching can also be used to make a solely rule-based decision about the similarity of a given case to the new situation. Based on a set of rules describing how similarity is computed, on a case-by-case basis, the system compares the selected features of the new situation to each case in the case-base. Through using the set of rules a score is computed for each case.

4.5 Section five

A "threshold" can be set for use during retrieval. A threshold tells the ESTEEM system what is an acceptable match to display to the end-user. The threshold is displayed as a percentage of similarity. Similarity is described to the end-user as a number from 0 to 100, with 0 being the weakest similarity measurement and 100 the strongest. The threshold provides a cut-off for displaying retrieved cases.

5. DISCUSSSION

Similes and metaphors could also allow us to partially solve the difficulty of dealing with context in spatial similarity. Spatial metaphors are not only relevant for human communication but also for objects organisation. People navigate in real life spatial environments based on very vague descriptions and facts. They are also familiar with memorising relationships of objects in spatial terms. Spatial representations are used frequently to convey one or more attributes of the information objects to the user: sorting, grouping and so forth. CBR allows us, coupled with a GIS to sort and catergorise spatial data according to their similarities. CBR allows us to change the similarity scores based on the context. In particular the similarity process of esteem highlights the flexibility of using such a system integrated with a GIS to make retrieving spatial similarities adaptable to different contexts. Are there any opportunities for similes and metaphors in geocomputation?

REFERENCES

Aamodt, A., and Plaza, E., 1994, "Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches." *Artificial Intelligence Communications,* 7 (1).

Benwell, G. L., Fryer, J., Whigham, P. A., Holt, A. (2000) "Towards a spatial similarity system for knowledge acquisition Aboriginal art and navigation sites in the Hunter Valley region, NSW, Australia." *The 12th Annual Colloquium of the Spatial Information Research Centre,* University of Otago, Dunedin, New Zealand. pp 281

Holt, A. and Benwell, G. L. (1996), "Case-Based Reasoning and Spatial Analysis", *Journal of the Urban and Regional Information Systems Association,* Vol. 8, No. 1, pp. 27-36

Holt, A. (1998), "The Matching and Ranking of Surface and Deep Similarities of Spatial Data", *Proceedings of the Spatial Information Research Centre's 10th Colloquium*, University of Otago, New Zealand, pp. 133-143.

Holt, A., (1999), "Spatial similarity and GIS: the grouping of spatial kinds." *The 11th Annual Colloquium of the Spatial Information Research Centre*, University of Otago, Dunedin, New Zealand, pp 241-250.

Holt, A. (2000), "Understanding environmental and geographical complexities through similarity matching" *Complexity International, 7:1-16,* (URL: http://www.csu.edu.au/ci/vol07/holt01/).

Holt, A. and Benwell, G. L. (1999), "Applying Case-Based Reasoning Techniques in GIS", *International Journal of Geographical Information Science*, 13(1), pp. 9-25.

Kaster, D.d.S. Medeiros C. B., da Rocha H. V.. "Aplicação de Raciocínio Baseado em Casos a Sistemas de Apoio à Decisão Ambiental." *GeoInfo 2000 II Workshop Brasileiro de GeoInformática*, Centro Anhembi, São Paulo, SP. http://www.tecgraf.puc-rio.br/geoinfo2000/anais.html

Kuhn, W. (1993). "Metaphors Create Theories for Users." In A. U. Frank and I. Campari (Eds.), *COSIT'93: Spatial Information Theory* (pp. 366-376). Springer-Verlag.

Macgillivray K., Wilson J. B. & Holt, A. (1999) "The Use of Artificial Intelligence techniques and GIS for predictive vegetation modelling." *The 11th Annual Colloquium of the Spatial Information Research Centre*, University of Otago, Dunedin, New Zealand, pp 305-306.

Ohlemüller, R., Holt, A. & Wilson, J. B. (2000) "Using a GIS/case-based reasoning hybrid for predictive mapping of potential natural forest vegetation in New Zealand." *The 4th International Conference on Integrating GIS and Environmental Modeling: Problems, Prospects and Research Needs*. Banff, Alberta, Canada. http://www.colorado.edu/research/cires/banff/upload/21/

Shi, X. and Yeh, A. G. O. (1999), "The Integration of Case-Based Systems and GIS in Development Control", *Environment and Planning B: Planning and Design,* 26, pp. 345-364.

Tversky, A. (1977) "Features of Similarity". *Psychological Review* 84(4): 327-352.

Tversky, A. & D. H. Krantz, (1970) "The dimensional representation and the metric structure of similarity data." *Journal of Mathematical Psychology*, Vol. 7. p572-597.

Yeh, A. G. O. and Shi, X. (1999), "Applying case-based reasoning to urban planning: a new planning-support system tool", *Environment and Planning B: Planning and Design*, Vol. 26, No. 1, pp. 101-115

Yeh, A. G. O. and Shi, X. (1998), "The Integration of Case-based Reasoning and GIS for Handling Planning Applications in Hong Kong", in Sikdar, P. K., Dhingra, S. L. and Krishna Rao, K. V. (eds.), *Computers in Urban Planning and Urban Management*, Vol. 2, New Delhi: Narosa Publishing House.