A Graph-based Alignment Approach to Similarity between Climbing Routes

Marc Wilkes and Krzysztof Janowicz

Institute for Geoinformatics, University of Muenster, Germany
marc.wilkes|janowicz@uni-muenster.de

Abstract. Reasoning about geographic features and feature types is a central functionality of spatial information retrieval and decision support systems. Measures of semantic similarity are such reasoning services which try to improve information retrieval by proposing similar features or types to a user's query. A major challenge for similarity theories is the alignment process, i.e., how to determine which characteristics describing features or their types are compared. Most theories disregard the internal structure of features and types, i.e., the spatio-temporal sequence of their characteristics, and simply assume that they can be modeled as unstructured bags of characteristics. In this paper, we demonstrate how to account for the spatial sequence of characteristics without the need to alter the similarity measures as such. To demonstrate our approach, a climbing route recommendation portal is introduced which proposes similar climbing routes based on the user's preferred routes.

1 Introduction and Motivation

With an increasing amount of user generated content on the Web the retrieval of relevant information gains additional importance [1]. Semantics-based information retrieval supports users beyond simple keyword-based search. For instance, similarity measures deliver rankings of contents which are similar to the user's query, and hence, support browsing through related content [2]. However, to deliver meaningful results requires an adequate representation of the searched contents. While products, such as televisions or computers can be described by key-value pairs, the representation of geographic features (and feature types) also requires the representation of the spatial (and in case of perdurants also temporal) sequence of these characteristics. Considering climbing routes, for instance, their similarity does not only depend on the rock formations and difficulty ratings of the compared pitches[1] but also on the distribution of these characteristics along the route. In psychology, this is known as alignment. Alignment-based models claim that similarity cannot be reduced to matching characteristics, but determining how these characteristics correspond to (*align with*) others [3, 4].

According to Janowicz et al. [2], the process of measuring semantic similarity can be divided into the following steps:

[1] A pitch is a section of rock between two belay points. A pitch's maximum length is the length of a climbing rope, which is usually 55 meters.

1. Selection of query and target features (or types).
2. Transformation of features (or types) to a normal form.
3. Definition of alignment matrices for feature (or type) characteristics.
4. Application of similarity functions for selected pairs of characteristics.
5. Determination of the standardized overall similarity.

In this work, we focus on the second and third step to account for the spatial sequence of feature characteristics without altering the actual similarity functions. To capture this sequence, a graph representation is proposed. As use case, we assume that a user queries a recommendation portal for climbing routes in order to retrieve a list of similar routes with respect to her preferred routes climbed before[2]. Whether the rock formations and difficulty ratings of two pitches are compared depends on whether they can be aligned, i.e., whether they occupy the same (or similar) position along the route. Vlachos et al. [5] examined the use of similarity for the retrieval of multidimensional trajectories. Whereas this approach accounts for the geometry, our approach is concerned with the sequence of characteristics independent of how they are represented (however, the recommendation portal uses description logics for representation).

2 Normalization and Alignment

In the following, the normalization and alignment process are discussed in detail using the climbing route use case.

Representation In order to allow for meaningful similarity measures, the spatial sequence of route characteristics is represented as depicted in figure 1. While the following representation focuses on climbing routes it can easily be adopted to other kinds of routes (e.g., bicycle routes).

A route R is a labeled, directed, and acyclic graph $R := (B, P)$ (called *topo graph* here), where B is a set of belays[3] represented as vertices of the graph, and P is a set of ordered pairs of belays, represented as edges of the graph and called pitches here. The order is given by the traversal direction, i.e., climbing routes are mastered from bottom to top. Each P has a set of characteristics $P := \langle L, D, RF \rangle$ assigned, where L is the climbing length, D the difficulty rated in UIAA scale[4], and RF is a pre-given set of rock formations, e.g., $\{arete, crack, overhang, wall, slab\}$. The representation of these characteristics is independent of the proposed alignment approach (however, we assume them to be assertions in a knowledge base). While climbing routes can be described by other characteristics as well, those discussed here are sufficient to demonstrate the taken approach.

[2] The service can be downloaded from http://sim-dl.sf.net/applications.

[3] A location where a climber provides protection to an ascending partner (in multi pitch climbs). In case of multi day bicycle routes, it could be a bed and breakfast place.

[4] http://theuiaa.org/guidebook_standards.html

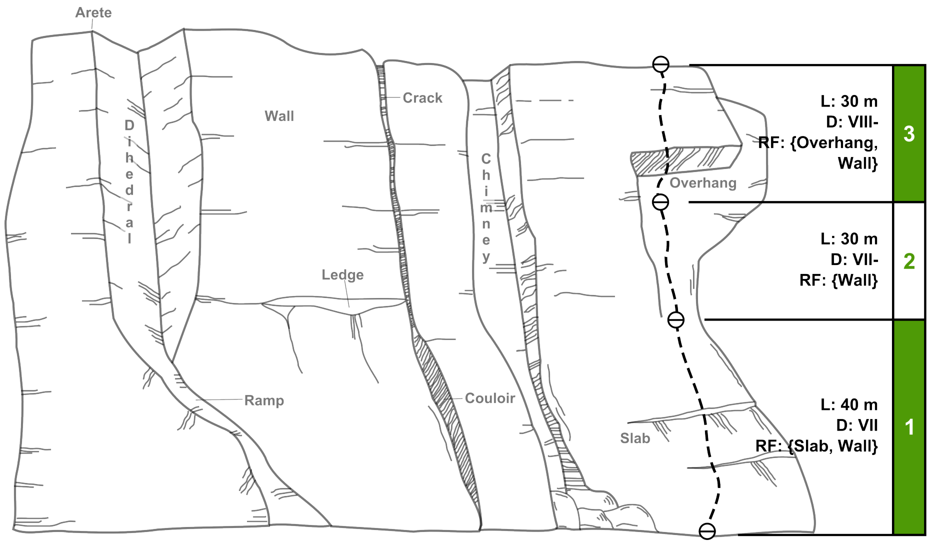


Fig. 1. From a climbing route feature to a *topo graph* (background graphic from [6]).

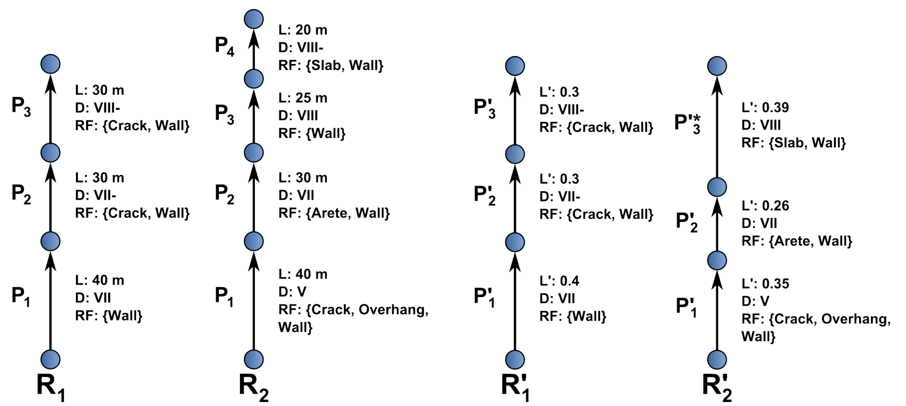
Normalization The alignment process ensures that only those characteristics that have a similar position within each route are selected for comparison. For example, if two routes both share the characteristic that they contain an overhang, where it occurs in the entry pitch of the first route and in the last pitch of the second route, this characteristic cannot be aligned [7]. Hence, sharing the characteristic of having an overhang does not necessarily increase the similarity between both routes. Before we can decide which route characteristics are aligned for similarity measurement, we have to ensure that both route representations (i.e., topo graphs) are normalized. This guarantees that both are described by the same number of pitches. Normalizing two topo graphs R_1 and R_2 involves two steps:

1. The total pitch length for both topo graphs R_1 and R_2 is standardized to 1.
2. If the topo graphs are described by a different number of pitches, neighboring pitches in the longer topo graph are merged until both contain the same number of pitches.

Since merging pitches means loosing information, this loss is minimized by merging those two neighboring pitches that are most similar and depends on the concrete similarity functions. When merging two standardized pitches P_1' and P_2' of R_1, rules have to define the characteristics of the new pitch $P_1'^*$. In our use case, the following rules are applied:

- $L'_{P_1'^*} = L'_{P_1'} + L'_{P_2'}$, where L' refers to the standardized pitch length
- $D_{P_1'^*} = max(D_{P_1'}, D_{P_2'})$
- $RF_{P_1'^*} = RF_{P_1'} \cup RF_{P_2'}$

An example for normalization is shown in figure 2. Figure 2(a) depicts the two topo graphs R_1 and R_2 before the normalization process. For reasons of simplification, we only consider the difficulty similarity as criterion for merging here. The two most similar (neighboring) pitches in R_2 are P_3 and P_4. The normalized graphs are shown in figure 2(b).



(a) Topo graphs R_1 and R_2. (b) Normalized topo graphs R'_1 and R'_2.

Fig. 2. Two topo graphs before and after the normalization process.

Alignment After normalization, the routes are ready for comparison. The alignment process determines those pairs of pitches that are considered within the overall route similarity by generating an alignment matrix. For lack of space, we assume that the alignment of characteristics only depends on the relative position of the pitch, while the implementation of the climbing route recommendation portal also considers other factors, such as the pitches' lengths.

The entries $m(i, j)$ of the alignment matrix are computed as defined in equation (1). $sim(P_{1i}, P_{2j})$ is a proxy for the currently applied similarity function, e.g., $sim_{\mathit{difficulty}}$, where P_{1i} is the i^{th} pitch of the route R_1 counted from the entry point of the route (and P_{2j} accordingly). While $sim(P_{1i}, P_{2j})$ measures the pitch similarity, the pitches' relative position within the routes is compared using ω_{pd} (equation (2)). It is defined as the relative pitch distance of the compared pitches to each other. The effect of ω_{pd} is that the *comparability* of two pitches P_{1i} and P_{2j} decreases as their relative distance increases[5].

$$m(i, j) = sim(P_{1i}, P_{2j}) * \omega_{pd}(P_{1i}, P_{2j}) \tag{1}$$

$$\omega_{pd}(P_{1i}, P_{2j}) = 1 - \frac{|i - j|}{\#\{P \mid P_i \in R_1\}} \tag{2}$$

[5] In fact, the length of pitches also needs to be taken into account for comparability. A description of how length information can be incorporated is given by Wilkes [8].

After the matrix is filled by computing the similarity and comparability between all pairs of pitches, the overall similarity can be determined (which is not discussed here). It is computed based on those pairs of pitches with the highest values in the matrix, where each pitch is selected exactly once.

The climbing route recommendation portal implementing this approach is shown in figure 3. Based on a set of user selected reference routes, the portal suggests a ranking of similar target routes.

Fig. 3. The user interface of the climbing route recommendation portal.

3 Conclusions

In this paper, we proposed how to account for the spatial sequence of characteristics of routes by enriching these characteristics with their relative position. This is achieved by modeling characteristics as labels of edges within a directed acyclic graph. Hence, our approach is independent of the used representation language of these characteristics, e.g., assertions in case of description logics (see SIM-DL [2]), and the applied similarity functions. Future work should address the integration of temporal aspects and additional topological relations (besides the *meet* relation [9]) of characteristics. Finally, the question whether characteristics can be aligned was determined based on similarity and comparability. Future work should investigate the relation between those factors (e.g., can comparability act as k.o. criterion).

Bibliography

[1] Rissland, E.L.: AI and similarity. IEEE Intelligent Systems **21**(3) (2006) 39–49

[2] Janowicz, K., Keßler, C., Schwarz, M., Wilkes, M., Panov, I., Espeter, M., Baeumer, B.: Algorithm, implementation and application of the SIM-DL similarity server. In Fonseca, F.T., Rodríguez, A., Levashkin, S., eds.: Second International Conference on GeoSpatial Semantics (GeoS 2007). Number 4853 in Lecture Notes in Computer Science, Mexico City, Mexico, Springer (November 2007) 128–145

[3] Goldstone, R., Son, J.: Similarity. In Holyoak, K., Morrison, R., eds.: Cambridge Handbook of Thinking and Reasoning. Cambridge University Press, Cambridge (2005)

[4] Li, B., Fonseca, F.: TDD - a comprehensive model for qualitative spatial similarity assessment. Spatial Cognition and Computation **6**(1) (2006) 31–62

[5] Vlachos, M., Kollios, G., Gunopulos, D.: Discovering similar multidimensional trajectories. Data Engineering, 2002. Proceedings. 18th International Conference on (2002) 673–684

[6] Lourens, T.: Klettern. Vom Einsteiger zum Könner. Delius Klasing Verlag (2007)

[7] Goldstone, R.: Similarity, interactive activation, and mapping. Journal of experimental psychology. Learning, memory, and cognition **20**(1) (1994) 3–28

[8] Wilkes, M.: A graph-based alignment approach to context-sensitive similarity between climbing routes (diploma thesis) (2008)

[9] Nedas, K., Egenhofer, M., Wilmsen, D.: Metric details of topological line–line relations. International Journal of Geographical Information Science **21**(1) (2007) 21–48