Metric space of hierarchies*

Sergey . Kirgizov @ LIP6.fr

July 10, 2013

Abstract

We turn a set of all hierarchies $\mathcal{H}$ of a graph $G = (V, E)$ into a metric space, using classical Hausdorff and Levenshtein distances.

Set of all hierarchies

Consider a graph $G = (V, E)$. Let $\mathcal{P}(V)$ be the set of all non-empty subsets of V. Denote by $\mathcal{H}$ the set of all possible hierarchies. Clearly, an hierarchy $H \in \mathcal{H}$ is a subset of $\mathcal{P}(V)$:

$$H \subset P(V) .$$

Suppose that one can define a metric between any two elements from $\mathcal{P}(V)$. In this case we can define the distance between two hierarchies $H_1, H_2 \in \mathcal{P}(V)$ as the Hausdorff distance between them.

Hausdorff distance

Given a metric space (M, d), define the distance between any point x from M and any non-empty subspace Y of M as the minimum of distances $d(x, y)$, along all y from Y:

$$\bar{d}(x, Y) = \min_{y \in Y} d(x, y) .$$

Define a distance between two subsets $X, Y \in M$ as the maximum of $\bar{d}(x, Y)$, taking over all $x \in X$:

$$\bar{\bar{d}}(X, Y) = \max_{x \in X} \bar{d}(x, Y)$$

But $\bar{\bar{d}}(X, Y)$ is not alwayse symmetric, thus the Hausdorff distance [1] is defined to be the maximum:

$$d_H(X, Y) = \max(\bar{\bar{d}}(X, Y), \bar{\bar{d}}(Y, X)) .$$

Note that the set of all non-empty subsets $\mathcal{P}(M)$ equiped with the Hausdorff distance d_H forms a metric space, when M is finite.

*Motivated by the talk “Évaluation et optimisation d’une partition hiérarchique de graphe” [3] performed by François Queyroi from LaBRI (Université Bordeaux I)

Levenshtein distance

Instead of writing a formula for the Levenshtein distance (also known as *edit distance*) we quote the Wikipedia article:

> The Levenshtein distance between two words is the minimum number of single-character edits (insertion, deletion, substitution) required to change one word into the other [2].

The $\mathcal{P}(V)$ equiped with the Levenshtein distance d_L forms a metric space.

Distance between two hierarchies

Consider two hierarchies H_1, H_2 as subspaces of the metric space $(\mathcal{P}(V), d_L)$, where d_L equals to the Levenshtein distance, we define the distance between hierarchies as:

$$d(H_1, H_2) = d_H(H_1, H_2) .$$

In this way we form a metric space $(\mathcal{H}, d_H)$.

References

[1] *Hausdorff distance.* `http://en.wikipedia.org/wiki/Hausdorff_distance`.

[2] *Levenshtein distance.* `http://en.wikipedia.org/wiki/Levenshtein_distance`.

[3] *Évaluation et optimisation d'une partition hiérarchique de graphe.* `http://www.complexnetworks.fr/evaluation-et-optimisation-dune-partition-hierarchique-de-graphe`.