

A New Class Of Proximity Data Obtained From Dictionary Networks

Camilo Garrido

Universidad de Chile, Santiago, Chile

Claudio Gutierrez

Universidad de Chile, Santiago, Chile

Guillermo Soto

Universidad de Chile, Santiago, Chile

Abstract

Background. Proximity data is a notion that indicates the degree of psychological closeness of concepts. It includes, among others, judgments of similarity, relatedness and cause-effect. Obtaining proximity data is challenging because it involves experts, corpora and people. On the other hand, dictionaries are fair representations made by experts (and thus, good proxies) of the lexicon and linguistic heritage of people.

Methods. We present a method to automatically obtain proximity data from dictionaries. We construct a network representation of a dictionary; exploit classical techniques on networks to build a similarity matrix; extract parameterized clouds of lexical proximity; test them with native speakers.

Results. Preliminary evaluations show that the method captures word associations significant to humans. Although the research was done in Spanish, the methods are easily reproducible in other languages.

Conclusions. Dictionaries are good sources of proximity data. We conjecture that dictionary networks are good proxies to human mind semantic associations.