Information Distance and Its Extensions⋆

Ming Li

School of Computer Science, University of Waterloo,
Waterloo, Ont. N2L 3G1, Canada
mli@uwaterloo.ca
http://www.cs.uwaterloo.ca/∼mli

Abstract. Consider, in the most general sense, the space of all information carrying objects: a book, an article, a name, a definition, a genome, a letter, an image, an email, a webpage, a Google query, an answer, a movie, a music score, a Facebook blog, a short message, or even an abstract concept. Over the past 20 years, we have been developing a general theory of information distance in this space and applications of this theory. The theory is object-independent and application-independent. The theory is also unique, in the sense that no other theory is "better". During the past 10 years, such a theory has found many applications. Recently we have introduced two extensions to this theory concerning multiple objects and irrelevant information. This expository article will focus on explaining the main ideas behind this theory, especially these recent extensions, and their applications. We will also discuss some very preliminary applications.

⋆ The full version of this paper is published in the Proceedings of the 14th International Conference on Discovery Science, Lecture Notes in Artificial Intelligence Vol. 6926.