

Enhancing feature selection with feature maximization metric

Jean-Charles Lamirel

Synalp Team, LORIA, Nancy, France

Email: lamirel@loria.fr

Abstract

This paper deals with a new feature selection and feature contrasting approach for classification of highly unbalanced textual data with a high degree of similarity between associated classes. The efficiency of the approach is illustrated by its capacity to enhance the classification of bibliographic references into a patent classification scheme. A complementary experiment is performed on a non textual dataset issued from the UCI repository.

Keywords: feature selection, feature maximization, classification, imbalanced data, similar classes.