



THE CHINESE UNIVERSITY OF HONG KONG
Department of Information Engineering

Seminar

**An Adjustment Method for Text Feature Weight
Based on Self-Adaptive Gaussian Filter Template**

by

Professor Yonghe Lu
School of Information Management
Sun Yat-Sen University
China

Date : 4 November, 2015 (Wednesday)
Time : 2:30 – 3:30pm
Venue : Room 1009 William M.W. Mong Engineering Building
The Chinese University of Hong Kong

Abstract

Traditional feature weight calculations do not consider the distribution of terms among classes or within a class so that the feature weight cannot accurately represent the meaning of texts. To overcome this problem, modified feature weight calculations have been widely used to improve the quality of text classifications. In spite of this, there still exist several challenges, such as only considering the distribution of terms within a class, or only considering the distribution of terms among classes, and without considering the impact of noise on features in text. In this paper, we propose an adjustment method based on self-adaptive Gaussian filter template (SAGFT) to solve the problem mentioned above. The new technique, SAGFT, includes a self-adaptive parameter that is trained with a corpus, and the self-adaptive parameter also is used to create new subsets within a class according to current texts. The self-adaptive Gaussian filter template is treated as an adjuster which can adjust text feature weights based on the distribution of text features in text space.

To demonstrate its achievement, SAGFT is tested separately with four feature selection algorithms, three methods of feature weighting, two methods of similarity calculation and three classifiers in R8 single-label corpora. Results suggest that the application of SAGFT increased the performance of the classifier. Based on the statistical studies, the new method clearly offers a greater performance than the method without using SAGFT in the majority of tested cases.

Biography

Yonghe Lu is a associate professor at Sun Yat-sen University in China, He is now a visiting scholar at the department of information engineering of CUHK . He received his Bachelor degree from Northeast Normal University in China in 1984. and his M.S. degree from Dalian Maritime University in 1996 in China. His current research interests include Semantic Analysis, Public Opinion Analysis, Text Mining, Intelligent Information Processing, Automatic Text Classification and Clustering. He published over 20 papers in peer-reviewed journals and in the international conferences in recent five years. He has presided 11 different level scientific research projects in recent five years, including the National Natural Science Foundation of China and the National High Technology Research and Development Program of China (863 Program).

**** ALL ARE WELCOME ****