

Title	Fタームによる特許分類のためのカーネル設計
Author(s)	三浦, 祥治
Citation	
Issue Date	2011-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/9670
Rights	
Description	Supervisor: Ho Bao TU, 知識科学研究科, 修士

Design of Kernel Functions for F-term-based Classification of Patents

Shoji Miura

School of Knowledge Science,
Japan Advanced Institute of Science and Technology
March 2011

Keywords: patent classification, kernel methods, multi label classification

Patents are rights for protecting intellectual and technical properties, products or inventions. The importance of the patent is increasing in the knowledge economy and society for strengthening the tactics of companies. As a national strategy, the Japanese government announced “Outline of Intellectual Property Strategy” in 2002. One of the strategies is to judge patents rapidly and accurately.

When people searches for new product ideas or inventions, submits a new patent application, or uses the patent for the enterprise tactics, people needs to retrieve relevant patents with the index systems; International Patent Classification (IPC), European Classification (ECLA), United States Patent Classification (USPC), File Index system (FI), or F-term system. IPC is the most major index system for classifying patent, ECLA is used in Europe, USPC is used in U.S.A, and FI and F-term is used in Japan. IPC is a hierarchical system, so IPC is explained about categories of intelligence or technologies of patent. F-term is multiple viewpoints and a hierarchical system, so F-term is explained about categories, materials, methods, operation, and so on. So when people search or filing to use the index system, F-term is more useful than IPC or FI. Because F-term can be search or survey patens and refined patents due to the multiple viewpoints.

In recent years, the number of applications of patents is huge. A special staff in Japan Patent Office (JPO) judges about 200 applications each year and the special staff is increasing the burden for judge a application, therefore, there are many applications to be read, understood, judge, patent administration annual report 2009 Edition. To address this issue,

JPO had hired 500 special staff from 2004 to 2009. In the result of addressing the issue, despite the huge number of applications each year there tends to decrease the number does not increase that is not processed. However, there is still a large number of unprocessed applications.

One of the answer is automatic labeling the index system by computer. Some researchers focused on IPC, some researchers focused on F-term. But, IPC can not narrow the patents fully. So, now researchers have studied automatic labeling to use F-term. One way of automatic labeling is classification. This is a technique of Machine Learning.

Support Vector Machines (SVMs) (Vapnik, 1998) are a widely used machine learning method which achieves good results in classification and regression tasks, including document classification, Li *et al*(2008). SVMs are good for binary classification, however, F-term classification is multi-labeled classification. One-vs.-rest strategy is used for multi classification. Previous works study what and how to use F-term speciality efficiently and solve the multi-labeled classification. They developed to SVMs for the specific features of the F-term patent classification task, Li *et al*(2008), Rikitoku (2007). These studies have got good results than previous studies. However, these studies did not focus on text information of the patent.

Patent is one of the text type. In the text classification, Kernel methods are achieved good result. So, the kernel methods can be adapted to patent classification and it is possible to improve the study of patent classification. Kernel methods is one of the SVMs algorithms, and the methods can deal with any type of data, e.g., structural data, graph data, or text data.

In this paper, we propose adapting a good model kernel method for patent classification. We select two kernel methods. One is linear kernel, and the other is RBF kernel. Both are popular, and good results. However, to get good results, it is necessary to set parameters of a kernel method. Purpose of this paper is we find best parameter of linear kernel and RBF kernel, and compare the performances of SVMs, linear kernel, and RBF kernel.

We have found the each best parameters. However, SVMs and RBF kernel could not classify F-term-based classification. Because total number of feature vectors is too big, however, number of feature vectors of each patent is less than total number of feature vectors, and dataset is imbalance. Experiments of linear kernel are same condition of others, however, the experiments get results that is not higher performance than previous work. From this results, there are efficient using linear kernel, and it is available to improve performance of F-term-based classification of patent.