論文 / 著書情報
Article / Book Information

題目(和文)	
Title(English)	Cluster Graph Classi cation Using the Generalized Shortest Path Kernel
著者(和文)	HermanssonLinusHakan
Author(English)	Linushakan Hermansson
出典(和文)	学位:博士(学術), 学位授与機関:東京工業大学, 報告番号:甲第10382号, 授与年月日:2016年12月31日, 学位の種別:課程博士, 審査員:渡辺 治,増原 英彦,鹿島 亮,鈴木 大慈,脇田 建
Citation(English)	Degree:Doctor (Academic), Conferring organization: Tokyo Institute of Technology, Report number:甲第10382号, Conferred date:2016/12/31, Degree Type:Course doctor, Examiner:,,,,
学位種別(和文)	博士論文
Category(English)	Doctoral Thesis
種別(和文)	論文要旨
Type(English)	Summary

（博士課程）
Doctoral Program

論 文 要 旨

THESIS SUMMARY

専攻：	Department of Mathematical and Computing Sciences 専攻	申請学位（専攻分野）： Academic Degree Requested	博士（学術） Doctor of Philosophy
学生氏名： Student's Name	Linus Hermansson	指導教員（主）： Academic Advisor(main)	渡辺　治
		指導教員（副）： Academic Advisor(sub)	

要旨（英文 800 語程度）
Thesis Summary（approx.800 English Words ）

Classifying graphs into different classes based on their structure is a problem which has become increasingly popular in recent years and that has many useful applications. Using various techniques to represent real world data as mathematical graphs and then classifying these graphs using different machine learning approaches; it has become possible to detect certain types of cancers, classify proteins by their enzyme class, which is very useful in bioinformatics, predicting weather, and much more.

Classifying graphs can be done in many different ways. The approach to classifying graphs taken in this thesis is one based on graph kernels and support vector machines, where we consider the supervised machine learning setting. This approach works as follows. We use different graph kernels to represent graphs as mathematical vectors. These vectors are then used to train a support vector machine to find a separating hyperplane which accurately classifies the vectors. Using this separating hyperplane, we are then able to predict the class of future graphs by transforming them into vectors using a graph kernel, and then testing which side of the separating hyperplane they lie. In recent years a large number of graphs kernels have been proposed, all with different weaknesses and strengths. Different graph kernels represent graphs as vectors by considering different properties of the graphs.

In this thesis we mainly investigate two different graph kernels, the well known and popular shortest path kernel and the recently published generalized shortest path kernel. The shortest path kernel compares graphs based on the number of node pairs which have a shortest path of the same length. Our new generalized shortest path kernel, just like the shortest path kernel, compares graphs based on the shortest path length, but it also compares graphs based on the number of shortest paths between node pairs. This means that the generalized shortest path kernel considers more information in a graph than the shortest path kernel. We compare these two graph kernels in the task of classifying graphs by the number of clusters they contain. By a cluster we mean a part of a graph which has a higher density of edges inside the cluster, as compared to between different clusters. Graphs with cluster structure are common in real world situations and are widely studied by researchers. A cluster in a graph can for instance represent a group of friends, in a social network graph.

The result of our work is an analysis of, in which situations, and why, the generalized shortest path kernel outperforms the shortest path kernel. In our analysis we are able to identify for which particular problems the generalized shortest path kernel can solve the problem with high accuracy and when it is not able to do so. We also provide a lot experimental results, which show that for a large number of datasets the generalized shortest path kernel outperforms the shortest path kernel. We provide several random models for generating random cluster graphs, which we use in our experiments in order to test the performance of the graph kernels.