

DOCTORAL THESIS

On feature selection, kernel learning and pairwise constraints for clustering analysis

Zeng, Hong

Date of Award:
2009

[Link to publication](#)

General rights

Copyright and intellectual property rights for the publications made accessible in HKBU Scholars are retained by the authors and/or other copyright owners. In addition to the restrictions prescribed by the Copyright Ordinance of Hong Kong, all users and readers must also observe the following terms of use:

- Users may download and print one copy of any publication from HKBU Scholars for the purpose of private study or research
- Users cannot further distribute the material or use it for any profit-making activity or commercial gain
- To share publications in HKBU Scholars with others, users are welcome to freely distribute the permanent URL assigned to the publication

On Feature Selection, Kernel Learning and Pairwise Constraints for Clustering Analysis

ZENG Hong

A thesis submitted in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

Supervisor: Dr. CHEUNG Yiu-ming

Hong Kong Baptist University

October 2009

Abstract

Given an objective function and a metric measure, the performance of most clustering algorithms strongly depend on the data representation. The first objective of this thesis is then to study how to learn a concise and informative representation for clustering analysis through feature selection and kernel learning. Besides, since the many clustering algorithms may not be able to produce user-desired groups, the second objective of this thesis is to investigate how to effectively incorporate user-provided pairwise constraints, which specify whether the pair of point should be in the same cluster or not, to assist the clustering.

Regarding to the first objective, we propose two feature selection methods as well as a kernel learning method. The first feature selection algorithm is proposed to reduce the dimension of data before performing clustering analysis. The importance of each feature is evaluated by its locality preserving capability, only those with higher locality preserving power are selected. To better characterize the local geometric structure, we formulate the locality preserving criterion based on the local kernel ridge regression method which is effective in modeling the local variation of target values. Experimental results on several benchmark datasets demonstrate that this method can maintain a low-dimensional representation for the data effectively. The second feature selection method is proposed to improve the performance of a specific clustering algorithm. This clustering algorithm employs the local learning idea to derive a clustering objective function which enforces the local smoothness on cluster labels. In the local learning, a local model is trained in the neighborhood of each sample, exclusively of samples out of the neighborhood. Therefore, it may suffer

from inconsistent estimation among adjacent neighborhoods, especially when there are insufficient number of samples in the neighborhood and there are many irrelevant features. To this end, we propose a new training paradigm in which local models are coupled by sharing a nonnegative feature weight vector among them. Such a paradigm is theoretically guaranteed not only be able to produce concise local model in each neighborhood, but also promote consistency among the local models. Thus the local smoothness criterion can be refined, leading to better partitioning. Furthermore, a kernel learning algorithm is derived from the second feature selection method using the kernel trick, where a convex combination of kernels is learned for the clustering. The effectiveness of the proposed approaches is validated on extensive real-world datasets.

For the second objective of this thesis, we propose to incorporate pairwise constraints into a clustering algorithm which finds maximum margin hyperplanes to separate the data. We further introduce a set of loss functions for effectively penalizing the violation of the given pairwise constraints. Unlike previous pairwise constrained clustering approaches which attempt to learn better distance metrics or improve the estimation for the underlying data distribution, the proposed approach aims at finding better clustering boundaries, thus requiring fewer model assumptions. The resulting problem is solved by the *constrained concave-convex procedure* (CCCP), with an efficient subgradient projection routine embedded. Comparison of the proposed method with several competitive approaches over a number of benchmark real-world datasets, shows that the proposed method is more computational efficient and able to propagate the pairwise constraints more effectively. Last but not the least, it also has good generalization performance on out-of-sample points.

Table of Contents

Declaration	i
Abstract	ii
Acknowledgements	iv
Table of Contents	v
List of Figures	viii
List of Tables	xi
Chapter 1 Introduction	1
1.1 Feature Selection for Clustering	2
1.2 Kernel Learning for Clustering	5
1.3 Pairwise Constrained Clustering	5
1.4 Overview of the Proposed Methods in the Thesis	8
1.5 Organization of the Thesis	10
Chapter 2 Surveys on Previous Works	11
2.1 Feature Selection for Clustering	11
2.2 Kernel Learning for Clustering	15
2.3 Pairwise Constrained Clustering	17
Chapter 3 Local Kernel Regression Score for Feature Selection	20
3.1 Introduction	20

3.2	Overview of the Kernel Ridge Regression and Neighborhood Graph	21
3.2.1	Kernel Ridge Regression	21
3.2.2	The Unsupervised and Supervised Neighborhood Graph	22
3.3	The Local Kernel Regression Score for Feature Selection	23
3.3.1	Relationship between the LKR Score and Laplacian Score	26
3.3.2	Time Complexity Analysis	28
3.4	Experimental Results	29
3.4.1	An Illustrative Example	30
3.4.2	Feature Selection for Clustering	31
3.4.3	Feature Selection for Classification	32
3.4.4	Studies of the Sensitivity of Parameters	33
3.5	Summary	34

Chapter 4 Feature Selection and Kernel Learning for Local Learning based Clustering 40

4.1	Introduction	41
4.2	Overview of the Local Learning Based Clustering Algorithm	42
4.3	Feature Selection for Local Learning Based Clustering	45
4.3.1	Update \mathbf{Y} for a Given $\boldsymbol{\tau}$	47
4.3.2	Update $\boldsymbol{\tau}$ for a Given \mathbf{Y}	48
4.3.3	The Complete Algorithm	50
4.4	Multiple Kernel Learning for Local Learning Based Clustering	51
4.4.1	Update \mathbf{Y} for a Given $\boldsymbol{\tau}$	52
4.4.2	Update $\boldsymbol{\tau}$ for a Given \mathbf{Y}	54
4.4.3	The Complete Algorithm	56
4.5	Discussion	57
4.5.1	Sparse Norm Equivalence	57
4.5.2	Relationship between LLC-fs and Q- α Algorithm	58
4.5.3	Relationship between LLC-fs and LLC-mkl	59
4.5.4	Time Complexity Analysis	60

4.6	Experimental Results	60
4.6.1	Experiments on Feature Selection	61
4.6.2	Experiments on Kernel Learning	70
4.7	Summary	77
Chapter 5 Pairwise Constrained Maximum Margin Clustering		79
5.1	Introduction	80
5.2	Related Works	80
5.3	Proposed Pairwise Constrained Maximum Margin Clustering	84
5.3.1	The Constrained Concave-convex Procedure for the Optimization	86
5.3.2	The Subgradient Projection Method for Quadratic Program .	90
5.3.3	Time Complexity Analysis	94
5.4	Experimental Results	95
5.5	Summary	101
Chapter 6 Conclusions and Future Work		106
6.1	Conclusions	106
6.2	Future Work	109
Bibliography		113
Appendices		128
Chapter A		128
A.1	Proof of Theorem 5.3.1	128
Curriculum Vitae		131