UNIVERSITI PUTRA MALAYSIA

ADAPTIVE SIMILARITY COMPONENT ANALYSIS IN NONPARAMETRIC DYNAMIC ENVIRONMENT

OMID SOJODISHIJANI

ITMA 2011 9

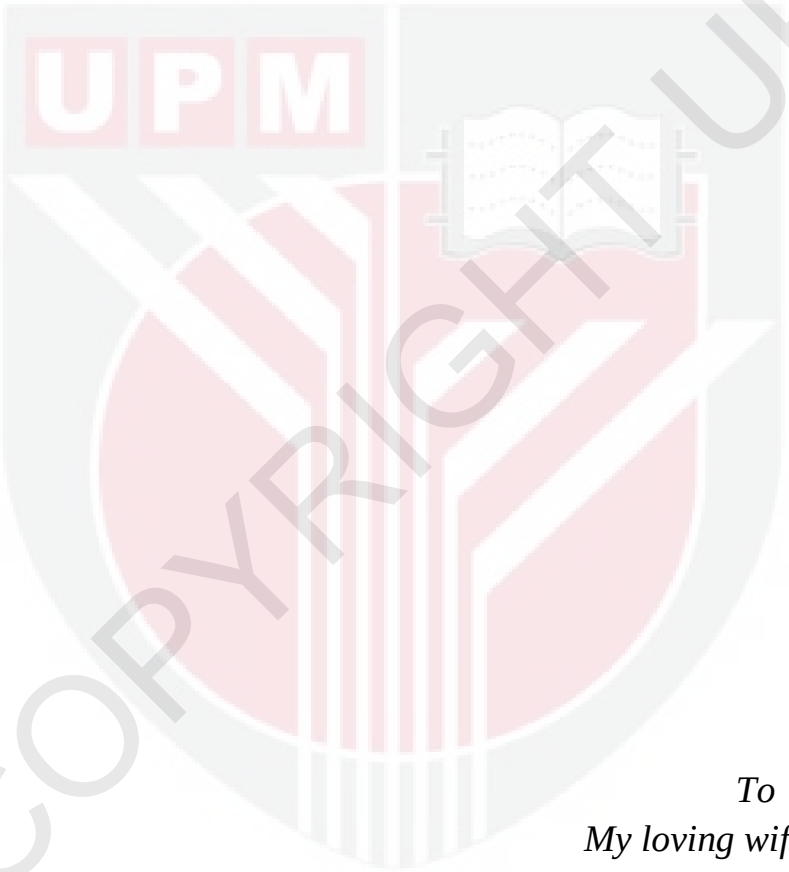
ADAPTIVE SIMILARITY COMPONENT ANALYSIS IN NONPARAMETRIC DYNAMIC ENVIRONMENT

By

OMID SOJODISHIJANI

Thesis Submitted to the School of Graduate Studies, Universiti Putra Malaysia, in Fulfilment of the Requirements for the Degree of Doctor of Philosophy

October 2011



To
My loving wife Nahid
and
My cute daughter Saba

Abstract of thesis presented to the Senate of Universiti Putra Malaysia in fulfilment of the requirement for the degree of PhD

ADAPTIVE SIMILARITY COMPONENT ANALYSIS IN NONPARAMETRIC DYNAMIC ENVIRONMENT

By

OMID SOJODISHIJANI

October 2011

Chairman: Associate Professor Abdul Rahman bin Ramli, PhD

Faculty: Institute of Advanced Technology

Pattern classification and recognition in low-rank distance metric dealing with nonparametric changes is an underlying problem in dynamic environment applications. Data arrives from operational field in a stream model and similarity-based classification algorithms must identify them with acceptable performance. Although, there are adaptive forms of independent feature extraction methods such as principle component analysis (PCA), linear discriminant analysis (LDA) and independent component analysis (ICA) to transform the training patterns to low dimensional space and/or improve the classifiers accuracy, they suffer from nonparametric changes in data over time.

This study is devoted to design a data-driven linear transformation to increase the performance of similarity-based classifiers in the presence of nonparametric changes of data over time. For this purpose, a nonparametric

multiclass component analysis technique in nonstationary environments is introduced. This generative model enables adaptive similarity-based classifiers to classify time-labeled inquiry pattern with superior accuracy in a low dimensional feature space.

In this thesis, an optimal transformation matrix is used to transform the time-labeled instances from original space to a new feature space in order to maximize the probability of selecting the correct class label for incoming instance by similarity-based classifiers. For this purpose, the most probable location of incoming instance for each class is estimated. Then, an optimal transformation matrix is computed by maximizing the information gain at the estimated points. By restricting the transformation matrix to a nonsquare matrix, the dimensions of feature space will be linearly reduced. Experimental results on real and synthesized datasets with real and artificial changes demonstrate the performance of the proposed method in terms of accuracy and dimension reduction in dynamic environments. In the case of real datasets, the proposed method yields 12.16% average misclassification error while the average misclassification error for five different methods GAM, TSY, NWKNN, DWM and FISH is 19.54%. Also, the results of experiments on synthesized datasets show that the proposed method yields 32.83% average misclassification error while average misclassification error of five different methods is 38.78%. From a dimensionality reduction evaluation aspect, the average misclassification error of the proposed method

in low-rank feature space is 9.6% and same error rate for three other well-known feature extraction methods is 21.21%.

The novelty of the proposed approach resides in the possibility to reduce the dimensions of feature space and simultaneously increase the accuracy of similarity-based classification method in an adaptive fashion in the nonparametric dynamic environment. Consequently, the proposed adaptive feature extraction technique and neighborhood-based classifier family are tightly integrated in an adaptive K-nearest neighbor classifier.

Abstrak tesis yang dikemukakan kepadaSenat Universiti Putra Malaysia sebagai memenuhi keperluan untuk ijazah Doktor Falsafah

ANALISIS KOMPONEN PERSAMAAN DALAM PERSEKITARAN DINAMIK

Oleh

OMID SOJODISHIJANI

Oktober 2011

Pengerusi: Profesor Madya Abdul Rahman bin Ramli, PhD

Fakulty: Institut Teknologi Maju

Pengelasan dan pengecaman corak pada peringkat rendah dalam metrik jarak yang melibatkan dengan perubahan maklumat bukan parameter adalah masalah asas dalam aplikasi persekitaran dinamik. Data yang tiba dari ruang operasi dalam model aliran dan algoritma pengelasan berasaskan kesamaan perlu mengenal pasti data ini dengan prestasi yang boleh diterima. Walaupun, terdapat beberapa penyesuai dalam kaedah pengekstrakan ciri bebas seperti analisis komponen prinsip (PCA), analisis pembeza linear (LDA) dan analisis komponen tak bersandar (ICA) untuk mengubah corak latihan kepada ruang dimensi rendah dan / atau meningkatkan ketepatan pengelas, tetapi persekitaran ini mengalami perubahan bukan parameter dalam data melalui perubahan masa.

Kajian ini menumpukan pada reka bentuk transformasi linear yang berasaskan data untuk meningkatkan prestasi pengelas berasaskan kesamaan dengan wujudnya perubahan data bukan parameter dari masa ke

masa. Untuk tujuan ini, satu teknik analisis komponen pelbagai kelas bukan parameter dalam persekitaran bukan pegun diperkenalkan. Model generatif (janaan) ini membolehkan pengelas penyesuaian berasaskan kesamaan mengelaskan corak siasatan berlabel masa dengan ketepatan yang unggul dalam ruang ciri dimensi rendah.

Suatu matriks transformasi optimum digunakan untuk mengubah keadaan berlabel masa dari ruang asal kepada ruang ciri baru untuk memaksimumkan kebarangkalian untuk memilih label kelas yang betul bagi keadaan yang masuk dengan menggunakan pengelas berasaskan persamaan. Untuk tujuan ini, lokasi yang paling mungkin pada keadaan yang masuk bagi setiap kelas dianggarkan. Selepas itu, satu matriks transformasi optimum dikira dengan memaksimumkan maklumat yang diperolehi pada titik anggaran. Dengan menghadkan matriks transformasi kepada matriks bukan segi-empat, dimensi ruang ciri dikurangkan secara linear. Keputusan eksperimen bagi set data sebenar dan yang disintesis dengan perubahan sebenar dan tiruan menunjukkan prestasi kaedah yang dicadangkan dari segi ketepatan dan pengurangan dimensi dalam persekitaran dinamik. Dalam kes set data sebenar, kaedah yang dicadangkan menghasilkan purata ralat kesalahan pengelasan adalah 12.16% manakala purata ralat kesalahan pengelasan bagi lima kaedah yang berbeza adalah 19.54%. Juga, keputusan eksperimen bagi set data yang disintesis menunjukkan bahawa kaedah yang dicadangkan menghasilkan 32.83% purata ralat kesalahan pengelasan manakala purata ralat kesalahan

pengelasan bagi lima kaedah yang berbeza adalah 38.78%. Dari aspek penilaian pengurangan dimensi, purata ralat kesalahan pengelasan bagi kaedah yang dicadangkan dalam ruang ciri berpangkat rendah adalah 9.6% dan kadar ralat yang sama bagi tiga kaedah pengekstrakan ciri terkenal lain adalah 21.21%.

Kebaharuan bagi pendekatan yang dicadangkan terdapat pada kebarangkalian untuk mengurangkan dimensi ruang ciri dan pada masa yang sama meningkatkan ketepatan pengelasan berasaskan kesamaan dalam satu fesyen penyesuaian tepat dalam masa untuk persekitaran dinamik bukan parameter. Oleh itu, teknik pengekstrakan ciri penyesuaian yang dicadangkan dan pengelas keluarga berasaskan kejiranan adalah padu rapat dalam satu pengelas penyesuai K-jiran terdekat yang tepat pada masa.

ACKNOWLEDGEMENTS

The journey has been challenging and exciting. My warm gratitude goes to the people who inspired me and helped in many ways. I especially thank my supervisor Associate Prof. Dr. Abd Rahman Ramli for introducing and showing the beauty of pattern recognition, for challenging discussions, advices and for the invaluable freedom I had in my research. I am deeply grateful to my co-supervisors Dr. Khairulmizam Samsudin and Associate Prof. Dr. Iqbal Saripan for their technical and content support. I am grateful to Professor Cesare Alippi and Dr. Manuel Roveri from Politecnico di Milano for their advices. I would like to thank Vahid Rostami for his support and friendship.

The long, hard process of completing a thesis would have been completely impossible without the support of many friends and colleagues at Institute of Advanced Technology and Qazvin Azad University. Most of all, I am grateful to my wife Nahid and my daughter Saba for their love and unfailing support. Their support and help is priceless to the completion of my study.

I certify that a Thesis Examination Committee has met on **24 October 2011** to conduct the final examination of Omid Sojodishijani on his thesis entitled "**Adaptive Similarity Component Analysis in Nonparametric Dynamic Environment**" in accordance with the Universities and University College Act 1971 and the Constitution of the Universiti Putra Malaysia [P.U.(A) 106] 15 March 1998. The Committee recommends that the student be awarded the PhD.

Members of the Thesis Examination Committee were as follows:

Azmi Zakaria, PhD
Professor
Faculty of Science
Universiti Putra Malaysia
(Chairman)

Mohammad Hamiruce Marhaban, PhD
Associate Professor
Faculty of Engineering
Universiti Putra Malaysia
(Internal Member)

Izhal b Abdul Halin, PhD
Senior Lecturer
Name of Faculty
Universiti Putra Malaysia
(Internal Member)

Kar-Ann Toh, PhD
Professor
School of Electrical and Electronic Engineering
Yonsei University
South Korea
(External Member)

_____

SEOW HENG FONG, PhD
Professor and Deputy Dean
School of Graduate Studies
Universiti Putra Malaysia

Date:

This thesis was submitted to the Senate of Universiti Putra Malaysia and has been accepted as fulfilment of the requirement for the degree of Doctor of Philosophy. The members of Supervisory Committee were as follows:

Abdul Rahman Ramli, PhD
Associate Professor
Faculty of Engineering
Universiti Putra Malaysia
(Chairman)

Khairulmizam Samsudin, PhD
Senior Lecturer
Faculty of Engineering
Universiti Putra Malaysia
(Member)

M. Iqbal Saripan, PhD
Associate Professor
Faculty of Engineering
Universiti Putra Malaysia
(Member)

BUJANG BIN KIM HUAT, PhD
Professor and Dean
School of Graduate Studies
Universiti Putra Malaysia

Date:

DECLARATION

I declare that the thesis is my original work except for quotations and citations which have been duly acknowledged. I also declare that it has not been previously, and is not concurrently, submitted for any other degree at Universiti Putra Malaysia or at any other institution.

_____

OMID SOJODISHIJANI

Date:

TABLE OF CONTENTS