A COMPARISON OF INCREMENTAL LEARNING IN ELECTROENCEPHALOGRAPHY (EEG) SIGNAL FOR PERSON AUTHENTICATION MODELLING

SOO PHENG KIAN

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

BORANG PENGESAHAN STATUS TESIS

JUDUL: A COMPARISON OF INCREMENTAL LEARNING IN ELECTROENCEPHALOGRAPHY (EEG) SIGNAL FOR PERSON AUTHENTICATION MODELLING

SESI PENGAJIAN: 2014/2015

Saya SOO PHENG KIAN

mengaku membenarkan tesis (PSM/~~Sarjana~~/~~Doktor Falsafah~~) ini disimpan di Perpustakaan Fakulti Teknologi Maklumat dan Komunikasi dengan syarat-syarat kegunaan seperti berikut:

1. Tesis dan projek adalah hak milik Universiti Teknikal Malaysia Melaka.
2. Perpustakaan Fakulti Teknologi Maklumat dan Komunikasi dibenarkan membuat salinan untuk tujuan pengajian sahaja.
3. Perpustakaan Fakulti Teknologi Maklumat dan Komunikasi dibenarkan membuat salinan tesis ini sebagai bahan pertukaran antara institusi pengajian tinggi.
4. ** Sila tandakan (/)

______	SULIT	(Mengandungi maklumat yang berdarjah keselamatan atau kepentingan Malaysia seperti yang termaktub di dalam AKTA RAHSIA RASMI 1972)
______	TERHAD	(Mengandungi maklumat TERHAD yang telah ditentukan oleh organisasi/badan di mana penyelidikan dijalankan)
______	TIDAK TERHAD	

(TANDATANGAN PENULIS)	(TANDATANGAN PENYELIA)
Alamat tetap: 40,Jalan Murni 17,Taman Malim Jaya,75250 Melaka.	PM. DR. CHOO YUN HUOY Nama Penyelia
Tarikh : ____________	Tarikh : ____________

CATATAN: * Tesis dimaksudkan sebagai Laporan Projek Sarjana Muda (PSM)
**Jika Tesis ini SULIT atau TERHAD, sila Lampirkan surat daripada pihak berkuasa.

A COMPARISON OF INCREMENTAL LEARNING IN ELECTROENCEPHALOGRAPHY (EEG) SIGNAL FOR PERSON AUTHENTICATION MODELLING

SOO PHENG KIAN

This report is submitted in partial fulfilment of the requirements for the Bachelor of Computer Science (Artificial Intelligence)

FACULTY OF INFORMATION AND COMMUNICATION TECHNOLOGY
UNIVERSITI TEKNIKAL MALAYSIA MELAKA
2015

DECLARATION

I hereby declare that this project report entitled

A COMPARISON OF INCREMENTAL LEARNING IN ELECTROENCEPHALOGRAPHY (EEG) SIGNAL FOR PERSON AUTHENTICATION MODELLING

is written by me and is my own effort and that no part has been plagiarized without citations.

STUDENT : ________________________ Date: ________________

(SOO PHENG KIAN)

SUPERVISOR : ________________________ Date: ________________

(PM. DR. CHOO YUN HUOY)

DEDICATION

To my beloved parents, Mr. Soo Mok Hua and Mrs. Tan Chew Suat, your love and support are my greatest inspiration upon accomplish this project.

To my dearest supervisor, PM. Dr. Choo Yun Huoy for being responsible, receptive and always by my side to encourage, guide and motivate me.

To my dear friends, especially Liew Siaw Hong, who help me to collection EEG dataset and provide some guidelines for this project. Besides that, I also want to thanks to Ku Man Yi and Ang Kuan Kee for your support and motivation throughout this project.

ACKNOWLEDGEMENTS

I would like to take this opportunity to express my gratitude to all those people who helped me in the completion of this final year project.

First of all, I would like to express my gratitude and appreciation to my dearest supervisor, Dr. Choo Yun Huoy who showed patience, tolerance, encouragement and her guidance throughout this project. I would also like to thank Liew Siaw Hong who has helped me in recording and understanding EEG signals dataset.

Furthermore, I would also like to thank my evaluator, Dr. Zeratul Izzah Binti Mohd Yusoh who has suggested in my project and helped me gain a better understanding in statistical result analysis.

Last but not least, I would like to express my deepest gratitude to my family members and friends for giving me support and encouragement throughout my project.

ABSTRACT

EEG applications commonly use small training data for analysis due to limited recording time. Besides, the consistency EEG signals of an individual can be affected by environmental factor or attention shift. Thus, incremental model is suitable for EEG analysis due to its capability of adaptation. However, there exists little research work focusing in this area especially on person authentication modelling. This project aims to compare the performance of the proposed Incremental Support Vector Machine, Incremental K-Nearest Neighbour and Hoeffding Tree for person authentication modelling. The experimental data involves VEP signals from 10 common human subjects recorded from using 10-20 system. Electrodes PO7, PO3, POZ, PO4, PO8, O1, OZ, O2 were used for recording EEG dataset. Feature extraction i.e. mean, coherence, cross-correlation, mutual information, wavelet packet decomposition (WPD) and hjorth parameter has been done on the recorded EEG dataset. The data were divided to 20 percent for training set while 80 percent for testing set. WEKA Knowledge Work Flow was used for incremental classification task for Incremental K-Nearest neighbour and Hoeffding Tree while Incremental Support Vector Machine was implemented in Matlab environment. The measurement of accuracy and true positive detection rate were used as the performance measure among for the comparison methods. Statistical tests i.e. the Shapiro-Wilk Normality test, Friedman Test and ANOVA test were used for validation purposes. From the statistical test and result analysis, Incremental Support Vector Machine showed the best performance among other models. This is because Incremental Support Vector Machine can handle EEG dataset with multi-class, polarity and many feature data. Incremental K-Nearest Neighbour and Hoeffding Tree proven equally good in the validation test. Nevertheless, hybrid Incremental Support Vector Machine model with Hoeffding Tree Model is suggested in the future work to overcome the shortcoming of Incremental Support Vector Machine in handling unbalanced class in person authentication framework.

ABSTRAK

Applikasi EEG biasanya menggunakan data latihan kecil untuk analisis kerana masa rakaman data terhad. Isyarat konsisten EEG individu boleh dipengaruhi oleh faktor persekitaran atau peralihan perhatian. Oleh itu, model tokokan sesuai untuk menganalisis EEG kerana keupayaan penyesuaian sendirinya. Kekurang penyelidikan memberi tumpuan dalam bidang pembelanjaran tokokan pada model pengesahan orang. Projek ini bertujuan untuk membanding prestasi tokokan Sokongan Mesin Vektor, tokokan K-Nearest Neighbour dan Hoeffding Tree dalam model orang pengesahan. Data eksperimen melibatkan isyarat VEP daripada 10 subjek manusia dan direkodkan mengguna 10-20 sistem. Elektrod PO7, PO3, POZ, PO4, PO8, O1, OZ, O2 digunakan untuk rakaman EEG dataset. Pengekstrakan ciri iaitu mean, kepaduan, cross-korelasi, mutual informasi, penguraian ombak paket (WPD) dan parameter Hjorth telah dilaksanakan ke atas EEG dataset. Dataset telah dibahagikan kepada 20 peratus bagi set latihan manakala 80 peratus untuk set ujian. WEKA Pengetahuan Aliran Kerja digunakan untuk tugas tokokan klasifikasi bagi Tokokan K-Nearest Neighbour dan Hoeffding Tree manakala Tokokan Sokongan Mesin Vektor dilaksanakan dalam persekitaran Matlab. Pengukuran ketepatan dan kadar positif benar telah digunakan sebagai ukuran prestasi antara kaedah tokokan. Ujian statistik iaitu ujian Shapiro-Wilk Normality, Friedman Ujian dan ujian ANOVA telah digunakan untuk tujuan pengesahan. Dari statistik ujian dan keputusan analisis, Tokokan Sokongan Mesin Vektor menunjukkan prestasi yang terbaik di kalangan model yang lain kerana Tokokan Sokongan Mesin Vektor boleh mengendalikan EEG dataset dengan pelbagai kelas, kekutuban dan data ciri yang banyak. Tambahan K-hampir Neighbour dan Hoeffding Tree terbukti sama baik dalam ujian pengesahan. Hibrid Tokokan Sokongan Mesin Vektor model pembelajaran dengan Hoeffding Tree Model dicadangkan dalam kerja masa depan untuk mengatasi kepincangan Tokokan Sokongan Mesin Vektor dalam mengendalikan tidak seimbang kelas dalam rangka pengesahan orang.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATONS

AI	-	Artificial Intelligence
ANN	-	Artificial Neural Network
ACC	-	Accuracy
AUC	-	Area under ROC curve
TPR	-	True Positive Rate
BCI	-	Brain Computer Interfaces
EEG	-	Electroencephalography
KNN	-	K-Nearest Neighbour
MLP	-	Multi-layered Perceptron
PIN	-	Personal Identification Number
ROC	-	Receiver Operating Characteristic
SVM	-	Support Vector Machine
VEP	-	Visual Evoked Potential
WEKA	-	Waikato Environment for Knowledge Analysis
IncSVM	-	Incremental Support Vector Machine
HT	-	Hoeffding Tree
ANOVA	-	Analysis of Variance

CHAPTER I

INTRODUCTION

1.1 Project Background

Electroencephalography (EEG) is a kind of signal. EEG is record in electrical mind signal. EEG is widely used recently. It is use in recording brain activity along the scalp. EEG measures voltage fluctuations from ionic current flows within the neurons of the brain. Such signals are usually below the noise level and thus not readily distinguished, so must use some methods and signal averaging to improve the signal-to-noise ratio.

Recently, researches in biometric security on person authentication. Other than traditional biological traits such as thumb print, new type of biometric traits that based on physiological signal such as EEG has been proposed. This is because EEG signal is unique and cannot be fake. Therefore, EEG signal can be used in person authentication. In person authentication, it require EEG dataset store in the database. When every time user use the system, user need to input signal to authenticate into system. The input signal will compare to the database signal which are generate an authentication system for the system.

Time by time, may be the signal of the person will have some changes due to environment factor or attention shift of the person. Thus, incremental learning of EEG signal is applied on the system. These will make the system have a more good performance and can learn time by time.

1.2 Problem Statement

From the real world application, we could not recording large dataset as the training data. Small dataset will be recorded first as the training data. As the new incoming testing data has been tested into the system, incremental learning will be carried out to gain the knowledge from the testing data and become parts of the training data. Few researches are focusing on noise created by attention shift in EEG analysis such as stress, fatigue and environment. Due to the changing of the EEG signal, it may have some method to let the dataset learning and updating old model time by time from the new data input into the system. Besides that, incremental learning is suggested in EEG signal for person authentication. There are also lack of research in comparing the incremental learning method in EEG signal for person authentication.

1.3 Objectives

The project embarks on the following objectives:

1. To compare incremental learning model for person authentication using EEG signal.
2. To design an incremental learning model to update knowledge by instance data stream.
3. To evaluate the proposed incremental learning model using classification approach

1.4 Scope

This project focus on the incremental learning method on person authentication using EEG signal. Besides, the training dataset consist 10 persons of the EEG signal to test on the incremental learning methods that are chosen and test for which methods has higher accuracy and true positive detection rate. The highest accuracy and true positive rate of the method will be the best method for incremental learning in EEG signal.

1.5 Project Significance

In this project, incremental learning methods is applied in system of EEG signal for person authentication will be performance faster and will get higher performance for the authentication system. System dataset do not need to retrain and just need to learn and keep updating from the incremental learning method.

1.6 Expected Output

An analysed result on accuracy and true positive rate and do the comparison on those selected incremental learning method. Recommendation can be made among the incremental learning method. Each incremental learning method should works in the EEG signal that are used for authentication method. Furthermore, knowledge work flow model of incremental KNN and Hoeffding Tree is designed in WEKA. Simple GUI implementation is done in Matlab for Incremental SVM for ease of training and testing EEG dataset with interface.

1.7 Summary

Electroencephalography (EEG) is a kind of signal widely use in variety of field recently. It is used in is the recording brain activity along the scalp. In this project, EEG signal is used in system authentication. But time by time, may be the signal of the person will have some changes. EEG data may change due to the emotional condition. Thus, incremental learning method of EEG signal is apply in the system. These will make the system learn time by time and have a better performance.

CHAPTER II

LITERATURE REVIEW AND PROJECT METHODOLOGY

2.1 Introduction

In this chapter, a literature review on person authentication in biometric security and various type of incremental learning data mining techniques in EEG signals have been studied. Person authentication in biometric security such as voice, retinal or iris scanning (Shedeed, 2011), fingerprint and face authentication algorithm are using in currently security technologies. In biometric person authentication, human behavioral characteristic and physiological which are unique, permanent and collectable can use for biometric person authentication (Dugelay et al., 2002).

While electroencephalogram (EEG) also can be used for biometric person authentication. EEG-based biometric person authentication is an emerging research topic and we believe that it may give a new research direction and use in application in the future. In the authentication system, the system consist of confirming and denying the identity request by a person (one-to-one matching). For the person authentication, it is more focusing on accept or reject a person claimed for identity. Person Authentication use the currently recorded biometric model by the system compare to the model in the database. Then differentiate the currently biometric model in the database. If the recorded biometric model is almost same as the biometric model in the database, the person grant the permission to access the system (Mill, 2007). Besides that, EEG base biometric security system also hard to fake and attack by the hacker. The biometric security system with fingerprint, voice and retina are not universal because the security system will be malfunction by the person dry skin, scars, loss of voice and etc. But, brain damage on the person is rarely occurred. Since that all human which are living has recordable EEG signal, so EEG feature is universal (Shedeed, 2011).

2.2 Incremental Learning

Incremental learning is a model of machine learning where the learning process take place time by time when new example has been added into system. Incremental learning are the online learning process (stream data) instance by instance base on the model had been learn from the batch learning (E.Utgoff, 2015). The most special things in incremental learning is we do not need a very sufficient training set to gain knowledge from the dataset, but the learning process will occur from the testing set. System will learning from time to time from the testing set without retrain by using whole dataset. From the history of the machine learning, a good training set that contain all necessary knowledge will be encourage and best for a system. Unfortunately, many real-world applications cannot fulfill the 'good' training set concept. It may affect by some distortion and noise when recording the training set. Besides that, some if the learning process may not possible to do easily or conveniently. This is because we need to have a longer time, larger storages, and cost to gain the sufficient large training dataset for the learning process. Even the training dataset are sufficiently obtained, the learning algorithm are hardly applied on to the dataset to all the training dataset because the training dataset are too large and cannot be loaded into computer memory (Geng and Smith-Miles, 2009). Therefore, incremental learning have been proposed to have a small training set at the beginning and learning from example from the testing set time to time.

Over the past decade of years, batch learning algorithm had been researched and investigated thoroughly. Batch learning algorithm are hard to apply in real world application because batch learning need to retrain of all the data when the new data comes. It is very time consuming and take a very long time to use in actual application (Guo et al., 2014). Besides that, real world Instead of retrain all the training dataset to gain the knowledge, it might be a better choice to choose incremental learning to learn time to time from the testing set using incremental learning algorithm. The learning example should be able to self-adapt from the changing environment (Geng and Smith-Miles, 2009).

Incremental learning is all about the learning approach of the classifier which can perform update knowledge base and assign new knowledge base to the previous knowledge base Below show the learning process in the most incremental algorithm In the incremental learning approach, the system with incremental learning can gain knowledge from new incoming data as well as evolve old training set become updated training set to be use on next training and testing process(Joshi, 2012). Below figure is the concept of incremental learning scenario block of data used to update the incremental classifier in the process if incremental learning over a period of time (Granger et al., 2008).

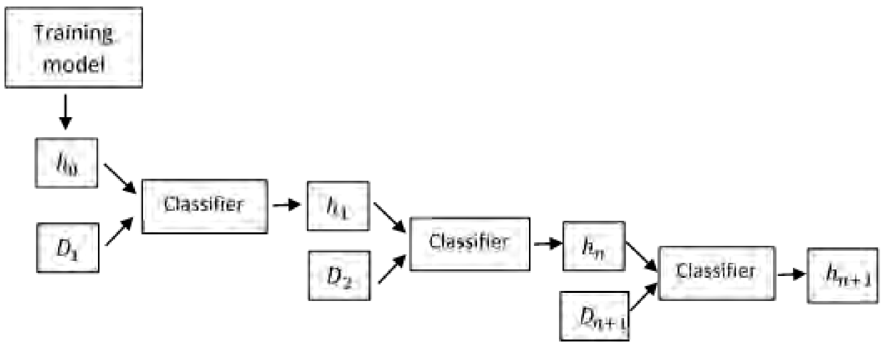

Figure 2.1: Incremental Learning Model

h_n = Hypothesis which build up by knowledge base/training model

D_n = New incoming testing data

2.3 Incremental Learning Techniques

Nowadays, dataset is very big. We need incremental learning algorithm to speed up the learning process from the data. Incremental learning has been applied in neural network based, SVM based, Nearest Neighbor based and tree based. There are no one implement incremental learning in naïve Bayes. Incremental learning acts as a main role in many real world application when new coming data is added into database instance by instance. There are some incremental learning technique had been applied in research. In the review paper of incremental learning technique, researcher has found out the different family of incremental classifier will have strength in certain application (Chen et al., 2009).

In the mathematical representation and algorithm, most of the incremental learning process is summarized in this algorithm:-

Let $U = \{ud_1, ud_2, ud_3, \dots ud_n\}$ be the new unlabeled data and
$L = \{ld_i : C_j \middle| i = 1\ to\ n, j = 1\ to\ m\}$.
Let Ic be the classifier that is used for incremental learning. Therefore,
$K = f\big(Ic(U_x), K_{prev}\big)\ where\ K = \{C_x, KB\}$

Figure 2.2: Mathematical Representation of Incremental Learning Process

The value of C_x can be existing class or new generate. K control the entire process. This is modeled and learned at every stage of incoming new data. The learning process is summarized in the figure of following algorithm:

1. For every $D_x | D_x$ Є U or L
2. Do

    Use KBprev

    If (D_x Є U)

    Classify $D_x, with\ f(IC)$

    Generate K

    Update $KBnew\ \leftarrow K + KBprev$

    Assign $KBprev\ \leftarrow KBnew$

Figure 2.3: Mathematical Incremental Learning Process with K control

2.3.1 Neural Base

Learning process in machine learning are uncertainty in many real scenarios. This challenges have open a new research to explore new algorithm that are able to handle with changes in the fundamental problem to be learnt (Perez-Sanchez et al., 2010). 2-layer feedforward neural network is an incremental learning algorithm with forgetting capability. But there are some strong point in the 2-layer feedforward neural network. It able to function in evolving environment. 2-layer feedforward neural network can reduce the memory requirement in the system processing incremental learning, maintain sufficient balance between new learning information and contain relevant old knowledge and fit dynamically on the forgetting capabilities (Perez-Sanchez et al., 2010)

While neural network also have its disadvantage. The trial-and-error design of the network is complex. The selection of the hidden nodes and training parameters is heuristic. Besides that, neural network for data mining is very heavy. Neural Network is data hunger which estimate the network weights requires large amounts of data, and this are very computer intensive (Cerny, 2010).

Multilayer perceptron (MLP) is a neural base algorithm has been propose in the research. In MLP, the approach for learning new knowledge from new incoming data will discard the old model of dataset and this scenario is called “catastrophic forgetting”. Catastrophic forgetting scenario may not suitable for some application because the original training data is no longer available (Polikar et al., 2001).

While ARTMAP algorithm is also another neural base incremental learning method. ARTMAP algorithm generated new decision clusters base on the new patterns that are different from previous instances. ARTMAP do not have scenario catastrophic forgetting. Besides that, ARTMAP can accommodate new classes without access to the previous seen data. From the research, ARTMAP are able to adapt in many different application. But ARTMAP is very sensitive noise data in the training data and generate a large number of clusters resulting in poor generalization performance due to over fit the training data (Polikar et al., 2001).