Improve Computer-Aided Arrhythmia Diagnosis with Deep Learning Techniques Using Undiagnosed Electrocardiography Samples

Yan Yan, *Student Member, IEEE*, and Lei Wang

***Abstract*—**

***Index Terms*—ECG Classification, Deep Learning, Arrythmia Analysis**

I. Introduction

THIS demo file is intended to serve as a "starter file" for IEEE journal papers produced under LaTeX using

In this paper, a new co-training style algorithm named Co-Forest, i.e. CO-trained random FOREST, is proposed. It extends the co-training paradigm by incorporating a well-known ensemble learning [13] algorithm named Random Forest [7] to tackle the problems of how to determine the most confident examples to label and how to produce the final hypothesis. Since ensemble learning has been successfully applied to many medical problems [35] [41] [42], the particular settings enables Co-Forest to exploit the power of ensemble for better performance of the learned hypothesis in semi-supervised learning. Since Co-Forest requires neither the data be described by sufficient and redundant attribute subsets nor special learning algorithms which frequently employ time-consuming cross validation in learning, it could be easily applied in CAD systems. Experiments on UCI data sets verify the effectiveness of the proposed algorithm. Case studies on three medical diagnosis tasks and a successful application to microcalcifi-

II. Dataset and Pre-processing

digm, which trains two classifiers through letting them label the unlabeled examples for each other. In co-training the data should be described by two sufficient and redundant attribute subsets, each of which is sufficient for learning and independent to the other given class label. Although co-training has already been successfully applied to some fields [4] [25] [30], the requirement on two sufficient and redundant attribute subsets is too strong to be met in many real-world applications. Goldman and Zhou [17] extended co-training by replacing the requirement on two sufficient and redundant attribute subsets with the requirement on two different supervised learning algorithms whose hypotheses partition the instance space into a set of equivalence classes. Ten-fold cross validation is frequently applied to find the confident examples to label in every training iteration and produce the final hypothesis, which makes both the learning process and prediction time-consuming. In this paper, a new co-training style algorithm named Co- Forest, i.e. CO-trained random FOREST, is proposed. It ex- tends the co-training paradigm by incorporating a well-known ensemble learning [13] algorithm named Random Forest [7] to tackle the problems of how to determine the most confident examples to label and how to produce the final hypothesis. Since ensemble learning has been successfully applied to many medical problems [35] [41] [42], the particular settings enables Co-Forest to exploit the power of ensemble for better perfor- mance of the learned hypothesis in semi-supervised learning. Since Co-Forest requires neither the data be described by sufficient and redundant attribute subsets nor special learning algorithms which frequently employ time-consuming cross validation in learning, it could be easily applied in CAD systems. Experiments on UCI data sets verify the effectiveness of the proposed algorithm. Case studies on three medical diagnosis tasks and a successful application to microcalcifi-digm, which trains two classifiers through letting them label the unlabeled examples for each other. In co-training the data should be described by two sufficient and redundant attribute subsets, each of which is sufficient for learning and independent to the other given class label. Although co-training has already been successfully applied to some fields [4] [25] [30], the requirement on two sufficient and redundant attribute subsets is too strong to be met in many real-world applications. Goldman and Zhou [17] extended co-training by replacing the requirement on two sufficient and redundant attribute subsets with the requirement on two different supervised learning algorithms whose hypotheses partition the instance space into a set of equivalence classes. Ten-fold cross validation is frequently applied to find the confident examples to label in every training iteration and produce the final hypothesis, which makes both the learning process and prediction time-consuming. In this paper, a new co-training style algorithm named Co- Forest, i.e. CO-trained random FOREST, is proposed. It ex- tends the co-training paradigm by incorporating a well-known ensemble learning [13] algorithm named Random Forest [7] to tackle the problems of how to determine the most confident examples to label and how to produce the final hypothesis. Since

Y. Yan is with the Shenzhen Key Laboratory for Low-cost Healthcare, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences. No. 1068, Xueyuan Road, Nanshan District, Shenzhen, Guangdong Province, China-mail: (yan.yan@siat.ac.cn).

L. Wang is with the Shenzhen Key Laboratory for Low-cost Healthcare, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences. No. 1068, Xueyuan Road, Nanshan District, Shenzhen, Guangdong Province, China.

Manuscript received ;

ensemble learning has been successfully applied to many medical problems [35] [41] [42], the particular settings enables Co-Forest to exploit the power of ensemble for better performance of the learned hypothesis in semi-supervised learning. Since Co-Forest requires neither the data be described by sufficient and redundant attribute subsets nor special learning algorithms which frequently employ time-consuming cross validation in learning, it could be easily applied in CAD systems. Experiments on UCI data sets verify the effectiveness of the proposed algorithm. Case studies on three medical diagnosis tasks and a successful application to microcalcifi-

III. Representations Learning with Deep Learning Methods

A. Sparse Autoencoder

digm, which trains two classifiers through letting them label the unlabeled examples for each other. In co-training the data should be described by two sufficient and redundant attribute subsets, each of which is sufficient for learning and independent to the other given class label. Although co-training has already been successfully applied to some fields [4] [25] [30], the requirement on two sufficient and redundant attribute subsets is too strong to be met in many real-world applications. Goldman and Zhou [17] extended co-training by replacing the requirement on two sufficient and redundant attribute subsets with the requirement on two different supervised learning algorithms whose hypotheses partition the instance space into a set of equivalence classes. Ten-fold cross validation is frequently applied to find the confident examples to label in every training iteration and produce the final hypothesis, which makes both the learning process and prediction time-consuming. In this paper, a new co-training style algorithm named Co- Forest, i.e. CO-trained random FOREST, is proposed. It ex- tends the co-training paradigm by incorporating a well-known ensemble learning [13] algorithm named Random Forest [7] to tackle the problems of how to determine the most confident examples to label and how to produce the final hypothesis. Since ensemble learning has been successfully applied to many medical problems [35] [41] [42], the particular settings enables Co-Forest to exploit the power of ensemble for better perfor- mance of the learned hypothesis in semi-supervised learning. Since Co-Forest requires neither the data be described by sufficient and redundant attribute subsets nor special learning algorithms which frequently employ time-consuming cross validation in learning, it could be easily applied in CAD systems. Experiments on UCI data sets verify the effectiveness of the proposed algorithm. Case studies on three medical diagnosis tasks and a successful application to microcalcifi-digm, which trains two classifiers through letting them label the unlabeled examples for each other. In co-training the data should be described by two sufficient and redundant attribute subsets, each of which is sufficient for learning and independent to the other given class label. Although co-training has already been successfully applied to some fields [4] [25] [30], the requirement on two sufficient and redundant attribute subsets is too strong to be met in many real-world applications. Goldman and Zhou [17] extended co-training by replacing the requirement on two sufficient and redundant attribute subsets with the requirement on two different supervised learning algorithms whose hypotheses partition the instance space into a set of equivalence classes. Ten-fold cross validation is frequently applied to find the confident examples to label in every training iteration and produce the final hypothesis, which makes both the learning process and prediction time-consuming. In this paper, a new co-training style algorithm named Co- Forest, i.e. CO-trained random FOREST, is proposed. It ex- tends the co-training paradigm by incorporating a well-known ensemble learning [13] algorithm named Random Forest [7] to tackle the problems of how to determine the most confident examples to label and how to produce the final hypothesis. Since ensemble learning has been successfully applied to many medical problems [35] [41] [42], the particular settings enables Co-Forest to exploit the power of ensemble for better perfor- mance of the learned hypothesis in semi-supervised learning. Since Co-Forest requires neither the data be described by sufficient and redundant attribute subsets nor special learning algorithms which frequently employ time-consuming cross validation in learning, it could be easily applied in CAD systems. Experiments on UCI data sets verify the effectiveness of the proposed algorithm. Case studies on three medical diagnosis tasks and a successful application to microcalcifi-

B. Restricted Boltzmann Machine

ch trains two classifiers through letting them label the unlabeled examples for each other. In co-training the data should be described by two sufficient and redundant attribute subsets, each of which is sufficient for learning and independent to the other given class label. Although co-training has already been successfully applied to some fields [4] [25] [30], the requirement on two sufficient and redundant attribute subsets is too strong to be met in many real-world applications. Goldman and Zhou [17] extended co-training by replacing the requirement on two sufficient and redundant attribute subsets with the requirement on two different supervised learning algorithms whose hypotheses partition the instance space into a set of equivalence classes. Ten-fold cross validation is frequently applied to find the confident examples to label in every training iteration and produce the final hypothesis, which makes both the learning process and prediction time-consuming. In this paper, a new co-training style algorithm named Co- Forest, i.e. CO-trained random FOREST, is proposed. It ex- tends the co-training paradigm by incorporating a well-known ensemble learning [13] algorithm named Random Forest [7] to tackle the problems of how to determine the most confident examples to label and how to produce the final hypothesis. Since ensemble learning has been successfully applied to many medical problems [35] [41] [42], the particular settings enables Co-Forest to exploit the power of ensemble for better perfor- mance of the learned hypothesis in semi-supervised learning. Since Co-Forest requires neither the data be described by sufficient and redundant attribute subsets nor special learning algorithms which frequently employ time-consuming cross validation in learning, it could be easily applied in CAD systems. Experiments on UCI data sets verify the effectiveness of the proposed

algorithm. Case studies on three medical diagnosis tasks and a successful application to microcalcifi-digm, which trains two classifiers through letting them label the unlabeled examples for each other. In co-training the data should be described by two sufficient and redundant attribute subsets, each of which is sufficient for learning and independent to the other given class label. Although co-training has already been successfully applied to some fields [4] [25] [30], the requirement on two sufficient and redundant attribute subsets is too strong to be met in many real-world applications. Goldman and Zhou [17] extended co-training by replacing the requirement on two sufficient and redundant attribute subsets with the requirement on two different supervised learning algorithms whose hypotheses partition the instance space into a set of equivalence classes. Ten-fold cross validation is frequently applied to find the confident examples to label in every training iteration and produce the final hypothesis, which makes both the learning process and prediction time-consuming. In this paper, a new co-training style algorithm named Co- Forest, i.e. CO-trained random FOREST, is proposed. It ex- tends the co-training paradigm by incorporating a well-known ensemble learning [13] algorithm named Random Forest [7] to tackle the problems of how to determine the most confident examples to label and how to produce the final hypothesis. Since ensemble learning has been successfully applied to many medical problems [35] [41] [42], the particular settings enables Co-Forest to exploit the power of ensemble for better performance of the learned hypothesis in semi-supervised learning. Since Co-Forest requires neither the data be described by sufficient and redundant attribute subsets nor special learning algorithms which frequently employ time-consuming cross validation in learning, it could be easily applied in CAD systems. Experiments on UCI data sets verify the effectiveness of the proposed algorithm. Case studies on three medical diagnosis tasks and a successful application to microcalcifi-

IV. Hybrid System with Deep Learning Features

V. Experiments

ch trains two classifiers through letting them label the unlabeled examples for each other. In co-training the data should be described by two sufficient and redundant attribute subsets, each of which is sufficient for learning and independent to the other given class label. Although co-training has already been successfully applied to some fields [4] [25] [30], the requirement on two sufficient and redundant attribute subsets is too strong to be met in many real-world applications. Goldman and Zhou [17] extended co-training by replacing the requirement on two sufficient and redundant attribute subsets with the requirement on two different supervised learning algorithms whose hypotheses partition the instance space into a set of equivalence classes. Ten-fold cross validation is frequently applied to find the confident examples to label in every training iteration and produce the final hypothesis, which makes both the learning process and prediction time-consuming. In this paper, a new co-training style algorithm named Co- Forest, i.e. CO-trained random FOREST, is proposed. It ex- tends the co-training paradigm by incorporating a well-known ensemble learning [13] algorithm named Random Forest [7] to tackle the problems of how to determine the most confident examples to label and how to produce the final hypothesis. Since ensemble learning has been successfully applied to many medical problems [35] [41] [42], the particular settings enables Co-Forest to exploit the power of ensemble for better perfor- mance of the learned hypothesis in semi-supervised learning. Since Co-Forest requires neither the data be described by sufficient and redundant attribute subsets nor special learning algorithms which frequently employ time-consuming cross validation in learning, it could be easily applied in CAD systems. Experiments on UCI data sets verify the effectiveness of the proposed algorithm. Case studies on three medical diagnosis tasks and a successful application to microcalcifi

VI. Conclusions

ch trains two classifiers through letting them label the unlabeled examples for each other. In co-training the data should be described by two sufficient and redundant attribute subsets, each of which is sufficient for learning and independent to the other given class label. Although co-training has already been successfully applied to some fields [4] [25] [30], the requirement on two sufficient and redundant attribute subsets is too strong to be met in many real-world applications. Goldman and Zhou [17] extended co-training by replacing the requirement on two sufficient and redundant attribute subsets with the requirement on two different supervised learning algorithms whose hypotheses partition the instance space into a set of equivalence classes. Ten-fold cross validation is frequently applied to find the confident examples to label in every training iteration and produce the final hypothesis, which makes both the learning process and prediction time-consuming. In this paper, a new co-training style algorithm named Co- Forest, i.e. CO-trained random FOREST, is proposed. It ex- tends the co-training paradigm by incorporating a well-known ensemble learning [13] algorithm named Random Forest [7] to tackle the problems of how to determine the most confident examples to label and how to produce the final hypothesis. Since ensemble learning has been successfully applied to many medical problems [35] [41] [42], the particular settings enables Co-Forest to exploit the power of ensemble for better perfor- mance of the learned hypothesis in semi-supervised learning. Since Co-Forest requires neither the data be described by sufficient and redundant attribute subsets nor special learning algorithms which frequently employ time-consuming cross validation in learning, it could be easily applied in CAD systems. Experiments on UCI data sets verify the effectiveness of the proposed algorithm. Case studies on three medical diagnosis tasks and a successful application to microcalcifi-

Appendix A
Proof of the First Zonklar Equation

Appendix one text goes here.

Appendix B

Appendix two text goes here.

ACKNOWLEDGMENT

The authors would like to thank...

PLACE PHOTO HERE

Yan Yan received the B.Eng. degree and the MSc in Instrument Engineering at the Harbin Institute of Technology in 2010 and 2012 respectively. He worked as research assistance at the Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences from 2012 to 2014. He is working on his PhD in Computer Science started from 2014. His research interests are digital signal processing, machine learning, and control system.

PLACE PHOTO HERE

Lei Wang received the B.Eng. degree in information and control engineering, and the Ph.D. degree in biomedical engineering from Xian Jiaotong University, Xian, China, in 1995 and 2000, respectively. He was with the University of Glasgow, Glasgow, U.K., and Imperial College London, London, U.K., from 2000 to 2008. He is currently a full professor with the Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China. He has published over 200 scientific papers, authored four book chapters, and holds 60 patents. His current research interests include body sensor network, digital signal processing, and biomedical engineering

John Doe Biography text here.

Jane Doe Biography text here.