南京航空航天大学

论文集

（二〇一〇年） 第43册

计算机科学与技术学院

（第2分册）

南京航空航天大学科技部编

二〇一一年五月

计算机科学与技术学院

计算机科学与技术学院2010年发表论文目录

序号	姓名	单位	职称	论文题目	刊名	发表时间	类别
80	顾彬 王建东 李涛	043 043 南信工	博士 教授 副	Ordinal-Class Core Vector Machine	Journal of Computer Science and Technology	2010年25卷4期	
81	任勇军 王建东 庄毅 王箭 徐大专	043 043 043 043	博士 教授 教授 教授 教	标准模型下基于身份的两方认证密钥协商协议	北京理工大学学报	2010年30卷2期	
82	任勇军 王建东 王箭 徐大专 庄毅	043 043 043 043 043	博士 教授 教授 教授 教授	标准模型下基于身份的认证密钥协商协议	计算机研究与发展	2010年47卷9期	
83	黄玉划 王箭 王建东	043 043 043	博士 教授 教授	离散数学教学策略研究	计算机教育	2010年18期	
84	宫大力 黄玉划 刘飞	043 043 043	硕士 副教授	RC4算法研究与改进	中国电子学会第十七届信息论学术年会论文集	2010	
85	赵路 庄毅 刘毅	043 043 043	硕士 教授 教授	A Refined Distributed Parallel Algorithm For The Eigenvalue Problem Of Large-scale Matrix	2010 3rd international Conference on Biomedical Engineering and Informatics	2010年8卷3期	
86	赵学建 庄毅	043 043	博士 教	无线传感器网络自适应功率控制策略	电子与信息学报	2010年32卷9期	
87	薛佟佟 庄毅 高阳 张锐恒	043 043 043 043	硕士 教授 硕士 硕	Markov编码模型在传感器网络加密算法中的应用	信息安全与技术	2010年1卷6期	
88	胡全舟 庄毅 刘佳	043 043 043	硕士 教授 讲师	A Long-running Transaction Model of Workflow	2010 3rd international Conference on Biomedical Engineering and Informatics	2010年8卷3期	
89	赵学建 庄毅	043 043	博士 教	无线传感器网络能量空洞问题环节策略分析	四川大学学报（工程科学版）	2010年42卷2期	
90	高阳 庄毅 殷科科 薛佟佟	043 043 043 043	硕士 教授 硕士 硕	An Improved Genetic Algorithm for Wireless Sensor Networks Localization	BIC-TA 2010	2010年1卷	
91	薛佟佟 庄毅 高阳 张锐恒 倪天权	043 043 043 043 043	硕士 教授 硕士 硕士 博士	A Novel Encryption Algorithm Based on Attractor computation for wireless sensor network	The IEEE Fifth International Conference on Bio-Inspired Computing: Theories and Applications	2010年1卷	
92	严新鑫 庄毅 刘彧 宋通	043 043 043 043	硕士 教授 硕士 硕	一种路由驱动的无线传感器网络密钥管理方案	信息安全与技术	2010年9月第7期	
93	薛羽 庄毅 倪天权	043 043 043	硕士 教授 博士	One Improved Genetic Algorithm Applied in the Problem of Dynamic Jam Resource Scheduling with Multi-objective and Multi-constraint	2010IEEE Fifth international conference on Bio-inspored Computing:Theories and Applications Volume-1	2010年1卷01期	

94	肖 笛 庄 毅 赵 路 倪天权 刘树锋	043 043 043 043 043	硕士 教授 硕士 博士 高级 工	一种低开销的海面模型	计算机工程与应用	2010.12	
95	冯爱民 刘学军 陈 斌	043 043 043	副 教授 副 教 授	结构大间隔单类分类器	山东大学学报（工学版）	2010年40卷2期	
96	皮德常 刘 军 秦小麟	043 043 043	教 授 本科 教	A Grey Prediction Approach to Forecasting Energy Demand in China	Energy Sources, Part A	2010年32卷16期	
97	王明涛 皮德常	043 043	硕士 教授	基于参考线段的轨道聚类算法研究	2010 Asia-Pacific Youth Conference on Communication Technology	2010	
98	姚明宇 皮德常	043 043	硕士 教授	Chinese text clustering algorithm based k-means	International Conference on Services Science, Management and Engineering(Volume	2010	
99	陶运信 皮德常	043 043	硕 士 教	一种快速移动对象轨道聚类算法	高技术通讯	2010年20卷1期	
100	向学敏 皮德常	043 043	硕士 教授	Trajectory Simplification and Classification for Moving Object with Road-Constraint	International Conference on Intelligent	2010	
101	曹子宁	043	教 授	Bisimulations for Open Processes in Higher Order π-Calculus	In Proceeding of TASE	2010	
102	曹子宁	043	教 授	Refinement Checking for Interface Automata With Z Notation	In Proceeding of SEKE 2010.USA.399-	2010	
103	曹子宁	043	教 授	Reducing Higher Order pi-Calculus to Spatial Logics	CORR abs/ 10112896	2010	
104	曹子宁	043	教 授	Temporal Logics and Model Checking Algorithms for ZIA	In Proceeding of SEDM	2010	
105	郑 青 曹子宁	043 043	硕士 教授	Process calculus with data structure and its model checking algorithm	In Proceeding of CCCM 2010.Yangzhou,China.	2010	
106	曹子宁	043	教授	Model Checking LOOP Programs	In Proceeding of SERP 2010.USA.92-99	2010	
107	崔新春 秦小麟	043 043	博 士 教	一种基于脆弱水印的可生存数据库篡改检测机制	计算机研究与发展（增刊）	2010年47卷	
108	涂金德 秦小麟 戴 华	043 043 043	本 科 教 授 博	基于双授权链集合的访问控制模型	计算机科学	2010年37卷7期	
109	蒋 鹏 秦小麟	043 043	博 士 教	一种动态场景中的视觉注意区域检测方法	小型微型计算机系统	2010年31卷4期	
110	朱广蔚 秦小麟 许 峰	043 043 043	博 士 教 授 副	一种基于负载均衡的多Agent路径规划算法	南京航空航天大学学报	2010年42卷2期	
111	刘亚丽 秦小麟 李博涵	043 043 043	博 士 教 授 博	Forward-Secure Blind Signature Schemes Based on the Variants of ElGamal	China Communications	2010年4卷	
112	刘 亮 秦小麟 刘亚丽	043 043 043	博 士 教 授 博	顽健的无线传感器网络K近邻查询处理算法	通信学报	2010年31卷11期	
113	刘宇雷 秦小麟 张 冰	043 043 043	博 士 教 授	基于R树的无线传感器网络节点管理技术	小型微型计算机系统	2010年31卷10期	
114	戴 华 秦小麟 李子玥 吴 冬	043 043 043 043	博 士 教 授 硕 士 硕	基于DBSUIM模型的数据库可以用户隔离技术	计算机研究与发展（增刊）	2010年47卷	
115	刘亚丽 秦小麟 殷新春 李博涵	043 043 043 043	博 士 教 授 硕 士	基于模m的n方根的前向安全数字签名方案的分析与改进	通信学报	2010年31卷6期	

116	陈逸菲 秦小麟	043 043	博士 教	一种路网中不确定移动对象范围查询分析方法	计算机研究与发展	2010年47卷6期	
117	陈逸菲 秦小麟 刘 亮	043 043 043	博士 教授 博	Uncertain Distance-Based Range Queries over Uncertain Moving Objects	Journal of Computer Science and Technology	2010年25卷5期	
118	戴 华 秦小麟 柏传杰	043 043 043	博士 教授	一种基于事务模板的恶意事务检测方法	计算机研究与发展	2010年47卷5期	
119	刘 亮 秦小麟 戴 华 严伟中 潘锦基	043 043 043 043 043	博士 教授 博士 博士 博	能量高效的无线传感器网络时空查询处理算法	电子学报	2010年38卷1期	
120	刘 亮 秦小麟 刘宇雷 李博涵	043 043 043 043	博士 教授 博士 博	链路感知的传感器网络空间范围查询处理算法	计算机研究与探索	2010年4卷8期	
121	徐丙凤 胡 军 黄志球 郭丽娟 张 剑	043 043 043 043 043	硕士 副教授 教授 硕	T-CBESD：一个构件化嵌入式软件设计模型验证工具	小型微型计算机系统	2010年31卷11期	
122	徐丙凤 胡 军 黄志球 郭丽娟 张 剑	043 043 043 043 043	硕士 副教授 教授 硕	构件化嵌入式软件模型非功能性质验证的工具实现	计算机科学	2010年37卷8期	
123	方元康 黄志球	043 043	博士 教授	Multi-objective fuzzy clustering method for image segmentation based on variable-length intelligent optimization algorithm	Lecture Notes in Computer Science	2010年6382期	
124	方元康 黄志球	043 043	博士 教授	A session identification algorithm based on frame page and pagethreshold	2010 3rd IEEE International Conference on Computer Science and Information Technology, ICCSIT 2010	2010.7.9-11	
125	方元康 黄志球	043 043	博士 教	An improved algorithm for session identification on web log	Lecture Notes in Computer Science	2010年6318期	
126	刘林源 黄志球 肖芳雄 沈国华	043 043 043 043	博士 教授 博士 副	Verification of Privacy Requirements in Web Services Composition	Second International Symposium on Data, Privacy, and E-Commerce	2010	
127	袁 敏 黄志球 曹子宁 肖芳雄	043 043 043 043	博士 教授 教授 博	一种扩充的π-演算及事务性等价关系研究	计算机研究与发展	2010年47卷3期	
128	张君华 黄志球 曹子宁	043 043 043	博士 教授 教授	Parallel programming patterns with granular computing	2010 International Conference on Computer Application and System Modeling	2010	
129	周 华 黄志球	043 043	硕士 教授	A Service-Centric Solution for Wireless Sensor Networks	PROCEEDINGS OF THE FIFTH INTERNATIONAL ICST CONFERENCE ON COMMUNICATION AND NETWORKING	2010年Ⅳ卷1期	
130	祝 义 黄志球 曹子宁	043 043 043	博士 教授 教	一种支持实时软件资源建模与分析的方法	东南大学学报（自然科学版）	2010年40卷3期	
131	祝义 黄志球 曹子宁	043 043 043	博士 教授 教	一种基于形式化规约生成软件体系结构模型的方法	软件学报	2010年21卷11期	

132	祝　义 黄志球 曹子宁	043 043 043	博士　教授　教授	An MDE Based Approach for Generating Software Architecture Models from Formal Specifications	The 10th International Conference on Quality Software	2010年11卷3期	
133	张君华 黄志球 曹子宁	043 043 043	博士　教授　教	Model Checking probabilistic timed automata in the presence of uncertainties	Journal of Computational Information Systems	2010年6卷7期	
134	洪　宏 黄志球 沈国华 钱　巨 刘春勇	043 043 043 043 043	硕士　教授　副教授　讲	支持软件可信评估的框架及其应用研究	计算机科学与探索	2011年5卷2期	
135	徐丙凤 黄志球 魏　欧	043 043 043	博士　教授　副教授	Making Architectural Decisions Based on Requirements:Analysis and Combination of Risk-Based and Quality Attribute-Based Methods	Symposia and Workshops on Ubiquitous, Autonomic and	2010	
136	袁　敏 黄志球 李　祥 闫　艳	043 043 043 043	博士　教授　硕士　硕	Towards a Formal Verification Approach for Business Process Coordination	IEEE International Conference on Web Services	2010	
137	刘亚萍 黄志球 祝　义	043 043 043	硕士　教授　博	基于元建模的实时系统模型转换方法研究	小型微型计算机系统	2011年31卷11期	
138	祝　义 黄志球	043 043	博士　教	一种支持实时软件时间建模的形式化方法	解放军理工大学学报（自然科学版）	2010年11卷3期	
139	闫　艳 黄　志球 袁　敏 沈国华	043 043 043 043	硕士　教授　博士　副	面向服务的多参与者协调事务建模方法	计算机科学与探索	2011	
140	朱梧槚	043	教　授	ELEMENTARY INFINITY-THE THIRD TYPE OF INFINITY BESIDES POTENTIAL INFINITY AND ACTUAL INFINITY	Proceedings of the 9th international FLINS conference	2010	
141	朱梧槚	043	教　授	THE RELATION OF OPPOSITION BETWEEN POTENTIAL INFINITY AND ACTUAL INFINITY	Proceedings of the 9th international FLINS conference	2010	
142	朱梧槚	043	教　授	Medium Logic and the Crises of Mathematic and Physics	Proceedings of the 9th international FLINS conference	2010	
143	王立松 秦小麟 丁秋林	043 043 043	教授　教授　教授	Modeling Access Control Resource Based on Process Algebra	International Journal of Computer Science & Network Security, VOL.10 No.3,March 2010	2010年10卷3期	
144	王立松 秦小麟 丁秋林	043 043 043	教授　教授　教授	A calculus with resource usage and consumption	2nd International Conference on Software Engineering and Data Mining,SEDM 2010	2010.6	

ı B, Wang JD, Li T. Ordinal-class core vector machine. JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY (4): 699–708 July 2010. DOI 10.1007/s11390-010-1054-y

)rdinal-Class Core Vector Machine

3in Gu[1] (顾　彬), Jian-Dong Wang[1] (王建东), and Tao Li[2] (李　涛)

Department of Computer Science and Engineering, Nanjing University of Aeronautics and Astronautics Nanjing 210016, China

[2]College of Electronic and Information Engineering, Nanjing University of Information Science and Technology Nanjing 210044, China

E-mail: jsgubin@163.com; aics@nuaa.edu.cn; lthnxx@126.com

Received April 5, 2009; revised January 18, 2010.

Abstract Ordinal regression is one of the most important tasks of relation learning, and several techniques based on support vector machines (SVMs) have also been proposed for tackling it, but the scalability aspect of these approaches to handle large datasets still needs much of exploration. In this paper, we will extend the recent proposed algorithm Core Vector Machine (CVM) to the ordinal-class data, and propose a new algorithm named as Ordinal-Class Core Vector Machine (OCVM). Similar with CVM, its asymptotic time complexity is linear with the number of training samples, while the space complexity is independent with the number of training samples. We also give some analysis for OCVM, which mainly includes two parts, the first one shows that OCVM can guarantee that the biases are unique and properly ordered under some situation; the second one illustrates the approximate convergence of the solution from the viewpoints of objective function and KKT conditions. Experiments on several synthetic and real world datasets demonstrate that OCVM scales well with the size of the dataset and can achieve comparable generalization performance with existing SVM implementations.

Keywords support vector machine, ordinal regression, ranking learning, core vector machine, minimum enclosing ball

1 Introduction

In conventional machine learning and data mining research, predictive learning has been a standard inductive learning, where different sub-problem formulations were identified, such as classification, metric regression, and ordinal regression. In the ordinal regression problems, the training samples are marked by a set of ranks, which exhibits an ordering among the different categories. In contrast to metric regression problems, these ranks are of finite types and the metric distances between the ranks are not defined; in contrast to classification problems, these ranks are also different from the labels of multiple classes due to the existence of the ordering information[1]. So to sum up, ordinal regression is a special task of predictive learning.

Although classification and metric regression problems have been thoroughly investigated in the literatures, the ordinal regression problems have not received nearly as much attention yet. Nonetheless, the applications of the ordinal regression frequently occur in domains where human-generated data plays an important role. Examples of these domains include information retrieval[2-3], collaborative filtering[4], medical sciences[5], and forecasting alert level of flight delays[6-7]. Especially, we take forecasting alert level of flight delays for example: according to the number of the delayed flights, the extent of flight delays in an airport can be divided into five ordinal levels, such as "severe", "high", "elevated", "guarded" and "low", which are also represented by five colors such as red, orange, yellow, green and blue respectively. So the investigation especially for ordinal regression will be very significant.

Ever since Vapnik's influential work in statistical learning theory[8], support vector machines (SVMs) have gained profound interest because of good generalization performance, there are also several approaches based on SVMs proposed to tackle ordinal regression problems. For example, Herbrich *et al.*[2] applied the principle of Structural Risk Minimization[8] to ordinal regression leading to a new distribution-independent learning algorithm based on a loss function between pairs of ranks. The main difficulty of the approach is that the problem size of the formulation is a quadratic function of the training data size. To overcome this issue, Shashua and Levin[4] generalized the support

Regular Paper

This work was supported by the National High-Tech Research and Development 863 Program of China under Grant No. 2006AA12A106.

vector formulation for ordinal regression by finding $q-1$ separating hyperplanes which would separate the training data into q ordered classes. This was done by modeling the ranks as the intervals on the real line. But there still exists a problem with this approach, which is that the ordinal inequalities on the thresholds $b_1 \leqslant b_2 \leqslant \cdots \leqslant b_{q-1}$ are not included in the formulation, it might result in disordered thresholds at the solution. This can be handled by introducing explicit constraints in the problem formulation that enforce the inequalities on the thresholds[1]. According to this, Chu and Keerthi[1] proposed a new formulation which considers the training samples from all the ranks to determine each threshold and gave the Sequential Minimal Optimization (SMO) algorithm for finding the solution of this formulation. Besides, Cardoso and Pinto da Costa[9] also proposed a data replication method and mapped it into support vector machines.

Although several approaches based on SVMs have been proposed to tackle the ordinal regression problems, the scalability aspect of these approaches to handle large datasets still needs much exploration. Recently, by reformulating SVM's quadratic programming as a minimum enclosing ball (MEB) problem, Tsang *et al.* applied an efficient $(1+\epsilon)$-approximation algorithm[10-11] to obtain a close-to-optimal SVM solution, which is the so-called core vector machine (CVM)[12]. CVM was first proposed to tackle one-class L2-SVM and two-class L2-SVM, which has an asymptotic time complexity that is linear with the number of training samples and a space complexity that is even independent of the number of training samples. Experimental results also demonstrate that the CVM is as accurate as other state-of-the-art SVM implementations, but is much faster and can handle much larger datasets than existing scale-up methods. Then, Asharaf *et al.*[13] extended CVM to multiclass classification problem. Although CVM is an effective method for handling large dataset in practical application, it can only be used with certain kernel functions and kernel methods. For example, the very popular support vector regression cannot be used with CVM. To overcome this problem, Tsang *et al.* introduced the center-constrained MEB problem and proposed the generalized CVM[14], which can be used with any linear/nonlinear kernel and more general quadratic programming formulations[15]. Soon afterwards, in order to make CVM not require any numerical solver, Tsang *et al.* proposed the simpler CVM with enclosing balls[16], that is the so-called ball vector machine. Inspired by CVM and MEB, Shevade and Chu[17] also presented the MEB formulations for support vector ordinal regression, but they still adopted the SMO algorithm[18], instead of using the CVM-like algorithm to solve the resulting optimization problem.

In this paper, we will extend the CVM algorith to the ordinal-class data and propose the Ordinal-Cla Core Vector Machine (OCVM). As mentioned abov its asymptotic time complexity is also linear with th number of training samples, while its space complexit is independent with the number of training samples.

The rest of this paper is organized as follows. Sectio 2 gives a short introduction to the MEB problem firs The OCVM is presented in Section 3, which mainly ir cludes two parts: formulation of ordinal-class CVM an $(1+\epsilon)$-approximation algorithm. Experimental result are presented in Section 4, and the last section give some concluding remarks.

2 Minimum Enclosing Ball Problem

Given a set of points $S = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_l\}$, where eac $\boldsymbol{x}_i \in R^d$, the minimum enclosing ball of S (denote as $MEB(S)$) is the smallest ball that contains all th points in S[12]. As shown in Fig.1, when the poir $\boldsymbol{c}^*$ is the center of $MEB(S)$ and the radius is R the minimum enclosing ball $MEB(S)$ can also be de noted as $B(\boldsymbol{c}^*, R^*)$. Let K be a kernel function wit the associated feature map $\varphi : \boldsymbol{x} \rightarrow \varphi(\boldsymbol{x})$. The $K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \langle\varphi(\boldsymbol{x}_i), \varphi(\boldsymbol{x}_j)\rangle$, where $\langle\cdot,\cdot\rangle$ denotes the ir ner product in a high dimensional reproducing kern Hilbert space (RKHS). Now the primal problem for th minimum enclosing ball in the RKHS can be state as[12]:

$$\min_{\boldsymbol{c},R} R^2 \quad \text{s.t.} \quad \|\varphi(\boldsymbol{x}_i) - \boldsymbol{c}\|^2 \leqslant R^2, \quad i = 1, \ldots, l. \quad (1$$

The corresponding dual is:

$$\min_{\boldsymbol{\alpha}} \sum_{i,j=1}^{l} \alpha_i \alpha_j K(\boldsymbol{x}_i, \boldsymbol{x}_j) - \sum_{i=1}^{l} \alpha_i K(\boldsymbol{x}_i, \boldsymbol{x}_i)$$
$$\text{s.t.} \quad \sum_{i=1}^{l} \alpha_i = 1, \quad \alpha_i \geqslant 0, \quad i = 1, \ldots, l \quad (2$$

where $\boldsymbol{\alpha}_i$ is the Lagrange multiplier.

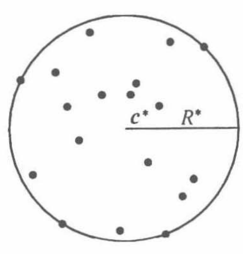

Fig.1. $MEB(S) = B(\boldsymbol{c}^*, R^*)$. Given a set of points S, $MEB($ is the minimum enclosing ball covering all the points of S w the center $\boldsymbol{c}^*$ and the radius R^*.

ɔw consider a situation where

$$K(\boldsymbol{x},\boldsymbol{x}) = \kappa \tag{3}$$

ıstant, and all the samples will be mapped on a rsphere in the RKHS. This restriction will cover kernel functions used in real-world applications, kample:

the isotropic kernel $K(\boldsymbol{x},\boldsymbol{y}) = \mathcal{K}(\|\boldsymbol{x}-\boldsymbol{y}\|)$, and Gaussian kernel is a special case of it;

the dot product kernel $K(\boldsymbol{x},\boldsymbol{y}) = \mathcal{K}(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y})$ (e.g., ıomial kernel) with normalized inputs;

any normalized kernel $K(\boldsymbol{x},\boldsymbol{y}) = \frac{\mathcal{K}(\boldsymbol{x},\boldsymbol{y})}{\text{c},\boldsymbol{x})\sqrt{\mathcal{K}(\boldsymbol{y},\boldsymbol{y})}}$.

ıder the restriction (3), the dual problem (2) can written as:

$$\begin{aligned} \min_{\boldsymbol{\alpha}} \ & \sum_{i,j=1}^{l} \boldsymbol{\alpha}_i \boldsymbol{\alpha}_j K(\boldsymbol{x}_i,\boldsymbol{x}_j) \\ \text{s.t.} \ & \sum_{i=1}^{l} \boldsymbol{\alpha}_i = 1,\ \boldsymbol{\alpha}_i \geqslant 0,\quad i=1,\ldots,l. \end{aligned} \tag{4}$$

whenever the kernel K satisfies the restriction ıny quadratic programming of the form (4) can be ded as an MEB problem (1).

$\boldsymbol{\alpha}^*$ is the optimal solution of problem (4), the pri- variables $\boldsymbol{c}^*$ and R^* of $B(\boldsymbol{c}^*, R^*)$ can be recovered lows:

$$\begin{aligned} \boldsymbol{c}^* &= \sum_{i=1}^{l} \boldsymbol{\alpha}_i^* \varphi(\boldsymbol{x}_i), \\ R^* &= \sqrt{\kappa - \sum_{i,j=1}^{l} \boldsymbol{\alpha}_i^* \boldsymbol{\alpha}_j^* K(\boldsymbol{x}_i,\boldsymbol{x}_j)}. \end{aligned} \tag{5}$$

ɾrdinal-Class Core Vector Machine

this section, we will first give the formulation of M, then present a $(1+\epsilon)$-approximation algorithm VM.

The Formulation of OCVM

dinal regression learning can be described as the ing: given an i.i.d. sample set $S = \{(\boldsymbol{x}_i, y_i)\}_{i=1}^{l} \sim$ and a mapping set $\mathcal{H} = \{h(\cdot) : \boldsymbol{X} \to Y\}$, a learn- ocedure selects one mapping h^l such that — using lefined loss — the risk function is minimized[2]. s statement, the input space $\boldsymbol{X} \subset \mathbb{R}^d$, the out- space $Y = \{r_1, \ldots, r_q\}$, which is ordered ranks $r_{q-1} \succ_Y \ldots \succ_Y r_1$, the number of k-th category training samples is denoted as l_k, and the total number of training samples $l = \sum_{k=1}^{q} l_k$.

As mentioned in the reduction framework of ordinal regression[19], the mapping function h consists of $q-1$ binary functions $f(\boldsymbol{x},k)$ with $k = 1,\ldots,q-1$. And if the mapping function $f(\boldsymbol{x},k)$ satisfies the condition of *rank-monotonic*, i.e., $f(\boldsymbol{x},1) \geqslant f(\boldsymbol{x},2) \geqslant \cdots \geqslant f(\boldsymbol{x},q-1)$, we will have that $h(\boldsymbol{x}) = 1 + \sum_{k=1}^{q-1}[f(\boldsymbol{x},k) > 0]$①. A simple *rank-monotonic* mapping function can be achieved by $q-1$ parallel discrimination hyperplanes in the RKHS, i.e., $f(\boldsymbol{x},k) = \langle \boldsymbol{w}, \varphi(\boldsymbol{x})\rangle + \boldsymbol{b}_k$ with $\boldsymbol{b}_1 \geqslant \cdots \geqslant \boldsymbol{b}_{q-1}$, which will be adopted in our proposed algorithm (see Fig.2).

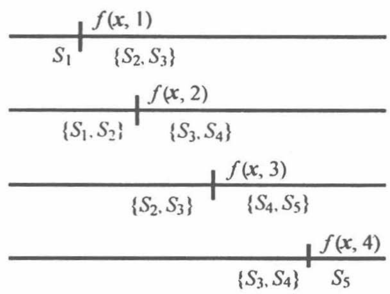

Fig.2. $q-1$ parallel discrimination hyperplanes. S_k denotes the set of k-th category training samples, and the k-th discrimination hyperplane is determined by s classes to its "left" and s classes to its "right". In this figure, $q = 5$, $s = 2$.

After defining the mapping set $\mathcal{H}$, we need to find a technique to minimize the risk function. According to the theory of generalization bounds of ordinal regression[19], we can bound the risk function by maximizing the minimum of the $q-1$ margins. Then the primal problem of OCVM can be presented as follows:

$$\begin{aligned} \min_{\boldsymbol{w},\boldsymbol{b},\rho,\boldsymbol{\varepsilon},\boldsymbol{\varepsilon}^*} \ & \|\boldsymbol{w}\|^2 + \sum_{k=1}^{q-1} \boldsymbol{b}_k^2 - 2\rho + \\ & C\sum_{k=1}^{q-1}\sum_{i=1}^{l_k}\sum_{j=in(k,s)}^{k} (\varepsilon_i^{k,j})^2 + \\ & C\sum_{k=1}^{q-1}\sum_{i=1}^{l_k}\sum_{j=k+1}^{su(k,s,q)} (\varepsilon_i^{*k,j})^2 \\ \text{s.t.} \ & \langle \boldsymbol{w}, \varphi(\boldsymbol{x}_i^j)\rangle \leqslant \boldsymbol{b}_k - \rho + \varepsilon_i^{k,j}, \\ & \text{for } j = in(k,s),\ldots,k \quad \text{and} \quad i = 1,\ldots,l_j; \\ & \langle \boldsymbol{w}, \varphi(\boldsymbol{x}_i^j)\rangle \geqslant \boldsymbol{b}_k + \rho - \varepsilon_i^{*k,j}, \\ & \text{for } \ j = k+1,\ldots,su(k,s,q) \quad \text{and} \\ & i = 1,\ldots,l_j; \end{aligned} \tag{6}$$

where $k = 1,\ldots,q-1$, s and C are user-defined

① The Boolean test $[\cdot]$ is 1 if the inner condition is true, and 0 otherwise.

parameters. The k-th discrimination hyperplane is determined by s classes to its "left" and s classes to its "right" with $1 \leqslant s \leqslant -1$ (see Fig.2); the scalars $\varepsilon_i^{k,j}$ and $\varepsilon_i^{*k,j}$ are relevant slack variables; $\rho/\|\boldsymbol{w}\|$ is so-called margin; two index functions $in(k,s)$ and $su(k,s,q)$ are defined respectively as follows:

$$in(k,s) = \max\{1, k-s+1\}$$
$$su(k,s,q) = \min\{q, k+s\}.$$

By the definition of the primal problem (6), the ordinal regression is reduced to $q-1$ binary classification problems. And if an original sample $(\boldsymbol{x}, y)$ is included in the k-th two-class training sample set, we will redefine it as $z = (\boldsymbol{x}, t, k)$, where $t = +1$ if $y > t$, and $t = -1$ otherwise. Thus, the original sample set S can be converted to $\widetilde{S} = \{\boldsymbol{z}_i = (\boldsymbol{x}_i, t_i, \phi(i))\}_{i=1}^{l'}$, where $\phi(i)$ means that z_i is included in the $\phi(i)$-th two-class training sample set.

Next, for the sake of presenting the dual function in a compact form, we will introduce the matrices $\boldsymbol{Q}$, $\boldsymbol{T}$ and $\boldsymbol{\Delta}$, which are all $l' \times l'$:

1) $\boldsymbol{Q}_{i_1,i_2} = \boldsymbol{t}_{i_1}\boldsymbol{t}_{i_2}K(\boldsymbol{x}_{i_1}, \boldsymbol{x}_{i_2})$.

2) $\boldsymbol{T}_{i_1,i_2} = \boldsymbol{t}_{i_1}\boldsymbol{t}_{i_2}$ if $\phi(i_1) = \phi(i_2)$, and $\boldsymbol{T}_{i_1,i_2} = 0$ otherwise.

3) $\boldsymbol{\Delta}_{i_1,i_2} = \frac{1}{C}$ if $i_1 = i_2$, and $\boldsymbol{\Delta}_{i_1,i_2} = 0$ otherwise; and let $\widetilde{\boldsymbol{Q}} = \boldsymbol{Q} + \boldsymbol{T} + \boldsymbol{\Delta}$, then the corresponding dual is:

$$\min_{\boldsymbol{\alpha}} \ \boldsymbol{\alpha}^{\mathrm{T}}\widetilde{\boldsymbol{Q}}\boldsymbol{\alpha}$$
$$\text{s.t.} \quad \boldsymbol{\alpha} \geqslant 0, \quad \boldsymbol{1}\cdot\boldsymbol{\alpha} = 1 \qquad (9)$$

where $\boldsymbol{\alpha}$ is the corresponding Lagrange multipliers.

Rewrite (9) in the form of (4) as:

$$\min_{\boldsymbol{\alpha}} \sum_{i,j=1}^{l'} \alpha_i\alpha_j\widetilde{K}(\boldsymbol{z}_i, \boldsymbol{z}_j)$$
$$\text{s.t.} \quad \sum_{i=1}^{l'}\alpha_i = 1, \quad \alpha_i \geqslant 0, \quad i = 1, \ldots, l' \qquad (10)$$

where $\widetilde{K}(\boldsymbol{z}_i, \boldsymbol{z}_j) = K(\boldsymbol{x}_i, \boldsymbol{x}_j) + T_{i,j} + \boldsymbol{\Delta}_{i,j}$. Since $K(\boldsymbol{x}, \boldsymbol{x}) = \kappa$, then $\widetilde{K}(\boldsymbol{z}, \boldsymbol{z}) = \kappa + 1 + \frac{1}{C} \stackrel{\text{def}}{=} \tilde{\kappa}$ satisfies the restriction (3), so the OCVM can be regarded as an MEB problem, in which φ is replaced by the nonlinear map $\tilde{\varphi}$ satisfying $\langle\tilde{\varphi}(\boldsymbol{z}), \tilde{\varphi}(\boldsymbol{z})\rangle = \widetilde{K}(\boldsymbol{z}, \boldsymbol{z})$. It can be easily verified that this $\tilde{\varphi}$ maps the training point $\boldsymbol{z}_i$ to a higher dimensional space as:

$$\tilde{\varphi}(\boldsymbol{z}_i) = \left[t_i\varphi(\boldsymbol{x}_i) \quad t_i\boldsymbol{\theta}_{\phi(i)} \quad \frac{1}{\sqrt{C}}\boldsymbol{e}_i \right]^{\mathrm{T}} \qquad (11)$$

where $\boldsymbol{\theta}_{\phi(i)}$ is a $(q-1)$-dimensional vector with all zeroes except that the $\phi(i)$-th position is equal to 1, and $\boldsymbol{e}_i$ is similarly the l'-dimensional vector with all zero except that the i-th position is equal to 1.

If $\boldsymbol{\alpha}^*$ is the optimal solution of problem (10), then

$$\boldsymbol{w} = \sum_{i=1}^{l'} t_i\boldsymbol{\alpha}_i^*\varphi(\boldsymbol{x}_i), \quad \langle\boldsymbol{w}, \varphi(\boldsymbol{x})\rangle = \sum_{i=1}^{l'} t_i\boldsymbol{\alpha}_i^*K(\boldsymbol{x}_i, \boldsymbol{x}) \qquad (1$$

and the primal variables $\boldsymbol{b}_k$ and ρ can also be recovere as:

$$\boldsymbol{b}_k = -\sum_{\forall i:\ \phi(i)=k} t_i\boldsymbol{\alpha}_i^*, \quad \rho = \sum_{i,j=1}^{l'} \boldsymbol{\alpha}_i^*\boldsymbol{\alpha}_j^*\widetilde{K}(\boldsymbol{z}_i, \boldsymbol{z}_j). \qquad (1$$

3.2 (1 + ϵ)-Approximation Algorithm of OCVM

Once the OCVM is formulated as an MEB prob lem, we get a modified RKHS with an associated kern function $\widetilde{K}$. This MEB problem in modified and RKH can be solved using the $(1+\epsilon)$-approximation algorith (Algorithm 1) introduced by Bădoiu and Clarkson[10 whose basic idea is to incrementally expand the core se $\widetilde{S}_t$ by including the sample that is the farthest from th center of the $MEB(\widetilde{S}_t)$ until the $(1+\epsilon)$-approximatio of $MEB(\widetilde{S}_t)$ covers all samples (see Fig.3).

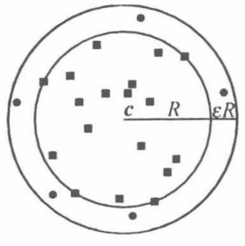

Fig.3. $B(\mathbf{c}, (1+\epsilon)R)$. The inner circle $B(\mathbf{c}, R)$ is the MEB of t set of squares and its $(1+\epsilon)$ expansion $B(\mathbf{c}, (1+\epsilon)R)$ (the out circle) covers all the points.

In Algorithm 1, the distance $\|\tilde{\varphi}(\boldsymbol{z}) - \boldsymbol{c}_k\|$ of a point from the center $\boldsymbol{c}_k$ of the $MEB(\widetilde{S}_k)$ in modified RKI can be computed as follows

$$\|\tilde{\varphi}(\boldsymbol{z}) - \boldsymbol{c}_k\| = \sum_{\boldsymbol{z}_i,\, \boldsymbol{z}_j \in \widetilde{S}_k} \alpha_i\alpha_j\widetilde{K}(\boldsymbol{z}_i, \boldsymbol{z}_j) - \sum_{\boldsymbol{z}_i \in \widetilde{S}_k} \alpha_i\widetilde{K}(\boldsymbol{z}_i, \boldsymbol{z}) + \widetilde{K}(\boldsymbol{z}, \boldsymbol{z}) \qquad (1$$

where $\boldsymbol{\alpha}_i$s are the Lagrange multipliers of finding t $MEB(\widetilde{S}_k)$ by (4).

According to the conclusion of [10], Algorithm 1 v find a $(1+\epsilon)$-approximation solution of the MEB at most $\frac{2}{\epsilon}$ iterations. In other words, the total nu ber of iterations τ is of $O(\frac{1}{\epsilon})$. And after finding

)-approximation solution of the MEB, the primal ɔles associated with the OCVM (i.e., weight $\boldsymbol{w}$, ɔ, slack errors $\boldsymbol{\varepsilon}$ and $\boldsymbol{\varepsilon}^*$) can be recovered from

$$[\boldsymbol{w} \quad \boldsymbol{b} \quad \sqrt{C}\boldsymbol{\varepsilon} \quad \sqrt{C}\boldsymbol{\varepsilon}^*]^{\mathrm{T}} = \boldsymbol{c}_\tau. \qquad (15)$$

lgorithm 1. $(1+\epsilon)$-Approximation Algorithm of 'M

: Initialize the core set $\widetilde{S}_0 = \{\widetilde{\varphi}(\boldsymbol{z}_0)\}$, the center $\boldsymbol{c}_0 = \widetilde{\varphi}(\boldsymbol{z}_0)$ and the radius $R_0 = 0$.

: Terminate if no $\widetilde{\varphi}(\boldsymbol{z})$ falls outside $B(\boldsymbol{c}_k, (1+\epsilon)R_k)$. Otherwise, find $\boldsymbol{z}_k$ such that $\widetilde{\varphi}(\boldsymbol{z}_k)$ is furthest away from $\boldsymbol{c}_k$. Set $\widetilde{S}_{k+1} = \widetilde{S}_k \cup \{\widetilde{\varphi}(\boldsymbol{z}_k)\}$.

: Find the new $MEB(\widetilde{S}_{k+1})$ according to (4).

: Increment k by 1 and go back to step 2.

ɩuppose that the time complexity for solving the $'B(\widetilde{S}_k)$ is of $O(|\widetilde{S}_k|^3)$, then for a given constant $\epsilon > 0$, time complexity of Algorithm 1 will be linear with number of training samples. As only one sample is led in the core set at each iteration, $|\widetilde{S}_k| = k+1$, core set initialization takes $O(1)$ time, the distance ıputations in step 2 take $O((k+1)^2 + (k+1)l') = O(k^2 + kl') = O(k^2 + kl)$ time, finding a new MEB in p 3 takes $O((k+1)^3) = O(k^3)$ time, and the other ɜrations take constant time. Hence, the k-th itera- ı takes a total of $O(kl + k^3)$ time, the total time ɛen by τ iterations is

$$\sum_{k=1}^{\tau} O(kl + k^3) = O(\tau^2 l + \tau^4) = O\Big(\frac{l}{\epsilon^2} + \frac{1}{\epsilon^4}\Big) \qquad (16)$$

ich is linear with l for a fixed ϵ.

Next, we consider the space complexity of Algo- ım 1. Suppose that the space complexity for solving $MEB(\widetilde{S}_k)$ is of $O(|\widetilde{S}_k|^2)$, then for a given constant $\cdot$ 0, the space complexity will be independent with number of training samples. As the training sam- s may be stored outside the core memory, the $O(l')$ ıce required will be ignored, hence the space com- xity for the k-th iteration is of $O(|\widetilde{S}_k|^2)$, the space nplexity for the whole procedure is of $O(\frac{1}{\epsilon^2})$, which ndependent of l for a fixed ϵ.

Experimental Results

Experiments are done with synthetic datasets and l world datasets, which demonstrate that OCVM les well with the size of the dataset and can achieve nparable generalization performance with existing M implementations.

All experiments are performed on Pentium-4 ma- nes having 1GB storage and running Windows XP. e value of ϵ is fixed at 10^{-3} in all the experiments ɩess otherwise specified.

4.1 Synthetic Datasets

To show the fast convergence of OCVM, we conducted the experiments using synthetic datasets. These synthetic datasets have five ordinal scale and each one was obtained from 2 multivariate normal distributions. The size of synthetic datasets ranges from 500 to 60 000. All experiments are done with the Gaussian kernel $K(\boldsymbol{x}, \boldsymbol{y}) = \exp(-\frac{\|\boldsymbol{x}-\boldsymbol{y}\|^2}{2\beta})$ with $\beta = 1000$, and with the parameters $C = 1000$, $s = 2$. From Fig.4, we can see that the numbers of iterations of the OCVM converge to 14, 7, 4 and 3 under the condition that $\epsilon = 0.05$, $\epsilon = 0.1$, $\epsilon = 0.2$ and $\epsilon = 0.3$ respectively, which verifies that OCVM will find an $(1+\epsilon)$-approximation solution in at most $\frac{2}{\epsilon}$ iterations. Fig.5 shows a comparison of training time for IMC(H),

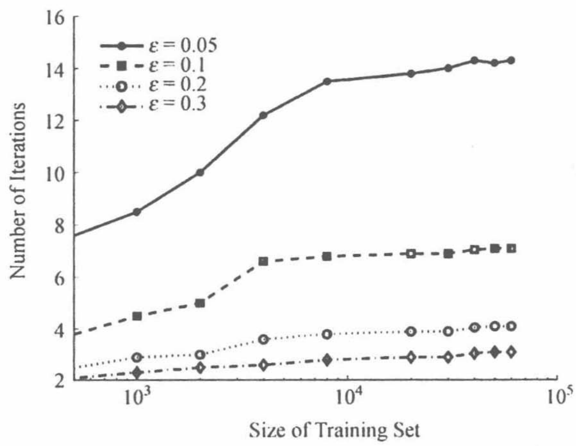

Fig.4. A plot of number of iterations for OCVM against the training set size (in log scale) with synthetic dataset, where the parameter $s = 2$.

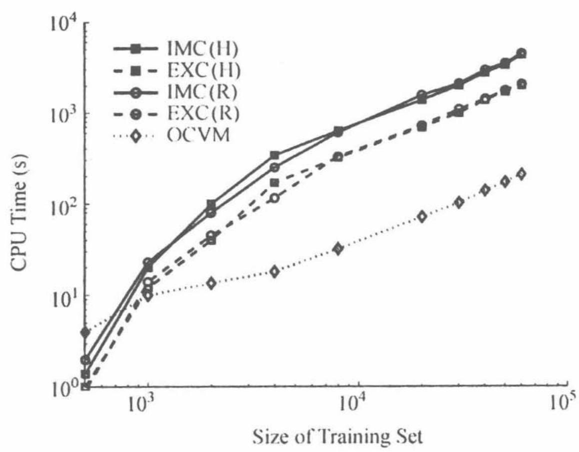

Fig.5. A comparison of training time (in seconds, in log scale) for IMC(H), EXC(H), IMC(R), EXC(R) and OCVM against the training set size (in log scale) with synthetic dataset, where the parameter $s = 2$.

EXC(H), IMC(R), EXC(R) and OCVM, where EXC(H) and IMC(H) are the abbreviations of the two SMO algorithms for the ordinal regression formulations with explicit constraints and implicit constraints on the thresholds②[1], EXC(R) and IMC(R) respectively represent the two approaches with explicit and implicit constraints on the radii[17]. From the figure, we can see that the time requirements of OCVM begin to exhibit a constant scaling with the training set size after processing around 4000 samples, and the increase of training time of OCVM is comparatively less than the ones of IMC(H), EXC(H), IMC(R) and EXC(R).

4.2 Real World Datasets

The OCVM algorithm is also evaluated on some real world datasets (i.e., the benchmark datasets for regression and the one for information retrieval) to show that it scales well with the size of the dataset and can achieve comparable generalization performance with existing SVM implementations.

In the experiments of comparing the generalization performance, we have utilized two evaluation metrics, which quantify the accuracy of predicted ordinal scales $\{\widehat{j}_1, \ldots, \widehat{j}_n\}$ with respect to true targets $\{j_1, \ldots, j_n\}$:

a) Mean absolute error: it is the average deviation of the prediction from the true target, i.e., $\frac{1}{n}\sum_{i=1}^{n}|\widehat{j}_i - j_i|$, in which we treat the ordinal scales as consecutive integers;

b) Mean zero-one error: it is simply the fraction of incorrect predictions, i.e., $\frac{1}{n}\sum_{i=1}^{n}[\widehat{j}_i \neq j_i]$③.

4.2.1 Benchmark Datasets for Regression

Table 1 summarizes the characteristics of the benchmark datasets for regression used in our experiments④, in which we randomly partition the datasets into the training set and test set, then select some as validation set. Originally, these datasets are used for metric regression problems. In order to make them used in ordinal regression problems, the target values are discretized into ten ordinal quantities using equal-frequency binning. The input vectors are normalized to zero mean and unit variance, coordinate-wise.

For the kernels, since our focus is on nonlinear kernels, we use the Gaussian kernel $K(\boldsymbol{x}, \boldsymbol{y}) = \exp(-\frac{\|\boldsymbol{x}-\boldsymbol{y}\|^2}{2\beta})$, where $\beta > 0$, in all the experiments, and a two-step grid search strategy with 5-fold cross validation is used to determine the optimal values of model parameters (the parameter β in the Gaussian kernel, and the regularization factor C) involved in the problem formulations: the initial search is dc on a 7×7 coarse grid linearly spaced in the regi $\{(\log_{10} C, \log_{10} \beta) | -3 \leqslant \log_{10} C \leqslant 3, -3 \leqslant \log_{10} \beta$ $3\}$, followed by a fine search on a 9×9 uniform grid l early spaced by 0.2 in the $(\log_{10} C, \log_{10} \beta)$ space. T validation set is specified in Table 1, and the test ror is obtained using the optimal model parameters f each formulation.

Table 1. Benchmark Datasets for Regression

Dataset	No. Attributes	No. Training Set	No. Validation Set	No. Test Set
Pyrimidines	27	50	20	24
MachineCPU	6	150	100	59
Boston	13	300	200	206
Abalone	8	1 000	1 000	3 177
Bank	32	3 000	1 500	5 192
Computer Activity	21	4 000	2 000	4 192
California	8	5 000	2 000	15 640
Census	16	6 000	4 000	16 784

In the experiments of comparing the generalizatic performance and studying the fast convergence, we ra domly partition each dataset into training/test spli as specified in Table 1. The partitioning is repeated times independently. And then we compare the gene alization capabilities and training time of our propose approach OCVM with the ones of EXC(H), IMC(H EXC(R) and IMC(R). The results are reported in T bles 2, 3 and 4 respectively. From these three table it is clear that our proposed approach scales well wi the size of the dataset and achieves comparable ge eralization performance as EXC(H), IMC(H), EXC(I and IMC(R).

Table 2. A Comparison of Training Time for the Five Algorithms Using a Gaussian Kernel (The parameter s of OCVM is set to 3. The targets of these benchmark datasets are discretized by 10 equal-frequency bins. The times are the averages over 20 trials.)

Training Set	CPU Time (s)				
	EXC(H)	IMC(H)	EXC(R)	IMC(R)	OCVM
Pyrimidines	0.06	0.12	0.03	0.10	0.10
MachineCPU	0.18	0.50	0.14	0.46	0.20
Boston	0.32	0.70	0.39	0.82	0.57
Abalone	11.84	22.71	13.15	19.33	8.98
Bank	106.70	201.93	143.19	295.23	16.74
Computer	149.10	356.54	193.67	389.41	19.34
California	178.27	412.17	223.76	498.46	24.85
Census	256.76	498.17	321.56	732.12	26.66

②The source code of IMC and EXC can be found at http://www.gatsby.ucl.ac.uk/~chuwei/svor.htm.

③The Boolean test [·] is 1 if the inner condition is true, and 0 otherwise.

④These regression datasets are available at http://www.liacc.up.pt/~ltorgo/Regression/Datasets.html.

Table 3. Mean Zero-One Errors of the Five Algorithms Using a Gaussian Kernel (The parameter s of OCVM is set to 3. The targets of these benchmark datasets are discretized by 10 equal-frequency bins. The results are the averages over 20 trials, along with the standard deviation.)

Dataset	EXC(H)	IMC(H)	EXC(R)	IMC(R)	OCVM
Pyrimidines	0.752 ± 0.063	0.719 ± 0.066	0.731 ± 0.062	0.729 ± 0.086	0.768 ± 0.068
MachineCPU	0.661 ± 0.056	0.655 ± 0.045	0.672 ± 0.046	0.658 ± 0.047	0.673 ± 0.054
Boston	0.569 ± 0.025	0.561 ± 0.026	0.575 ± 0.027	0.564 ± 0.025	0.568 ± 0.026
Abalone	0.736 ± 0.011	0.732 ± 0.007	0.724 ± 0.013	0.733 ± 0.009	0.737 ± 0.010
Bank	0.744 ± 0.005	0.751 ± 0.005	0.749 ± 0.006	0.752 ± 0.006	0.749 ± 0.006
Computer	0.462 ± 0.005	0.473 ± 0.005	0.470 ± 0.007	0.474 ± 0.006	0.468 ± 0.006
California	0.640 ± 0.003	0.639 ± 0.003	0.646 ± 0.004	0.640 ± 0.004	0.642 ± 0.004
Census	0.699 ± 0.002	0.705 ± 0.002	0.702 ± 0.003	0.706 ± 0.003	0.703 ± 0.002

Table 4. Mean Absolute Errors of the Five Algorithms Using a Gaussian K (The parameter s of OCVM is set to 3. The targets of these benchmark datasets are discretized by 10 equal-frequency bins. The results are the averages over 20 trials, along with the standard deviation.)

Dataset	EXC(H)	IMC(H)	EXC(R)	IMC(R)	OCVM
Pyrimidines	1.331 ± 0.193	1.294 ± 0.204	1.330 ± 0.194	1.290 ± 0.202	1.330 ± 0.198
MachineCPU	0.986 ± 0.127	0.990 ± 0.115	0.985 ± 0.128	0.989 ± 0.123	0.994 ± 0.126
Boston	0.773 ± 0.049	0.747 ± 0.049	0.779 ± 0.054	0.750 ± 0.051	0.756 ± 0.054
Abalone	1.391 ± 0.021	1.361 ± 0.013	1.401 ± 0.022	1.3581 ± 0.019	1.356 ± 0.024
Bank	1.512 ± 0.017	1.393 ± 0.011	1.511 ± 0.018	1.342 ± 0.013	1.501 ± 0.016
Computer	0.602 ± 0.009	0.596 ± 0.008	0.613 ± 0.011	0.599 ± 0.008	0.611 ± 0.010
California	1.068 ± 0.005	1.008 ± 0.005	1.072 ± 0.007	1.010 ± 0.008	1.007 ± 0.006
Census	1.270 ± 0.007	1.205 ± 0.007	1.268 ± 0.009	1.215 ± 0.008	1.283 ± 0.008

4.2.2 Benchmark Dataset for Information Retrieval

Ranking learning arises frequently in information retrieval. Liu *et al.*[20] built a benchmark dataset named LETOR⑤, which consists of 69623 references and 9999 queries with their respective ranked results. The relevance level of the references with respect to the given textual query were assessed by human experts, using a three rank scale: definitely, possibly, or not relevant.

Table 5. A Comparison of Training Time for the Five Algorithms Using a Linear Kernel (The parameter s of OCVM is set to 1. The times are the averages over 50 trials.)

No. Training Set	CPU Time (s)				
	EXC(H)	IMC(H)	EXC(R)	IMC(R)	OCVM
5 000	411	434	278	297	21
10 000	711	803	689	792	33
15 000	1 118	1 354	1 056	1 370	56
20 000	1 489	1 693	1 434	1 596	75

We randomly selected a subset from the whole database (with size chosen from {5 000, 10 000, 15 000, 20 000}) for training and then tested on the remaining references. For each size, the random selection was repeated 50 times. The training time and generalization performance of OCVM was compared against EXC(H), IMC(H), EXC(R) and IMC(R). The linear kernel $K(x_i, x_j) = \langle x_i, x_j \rangle$ was employed for all the five algorithms (especially, OCVM use the normalized linear kernel), the parameter s of OCVM is set 1. The training times are reported in Table 5, and the results of generalization performance are presented as boxplots in Fig.6. In this case, OCVM scales well with the size of the dataset and achieves comparable generalization performance as EXC(H), IMC(H), EXC(R) and IMC(R).

4 Conclusions

A scalable kernel method for ordinal regression, namely Ordinal-Class Core Vector Machine, is proposed in this paper. The proposed method can hurdle the large sample problem for ordinal regression effectively, because the theoretical analysis and experiments show that the method scales well with the size of the dataset and can achieve comparable generalization performance with existing SVM implementations. At last, some properties of OCVM are summarized as follows.

a) The method scales well with the size of the dataset and has the comparable generalization performance with existing SVM implementations;

⑤The dataset is available at http://research.microsoft.com/en-us/um/beijing/projects/letor/.

Bin Gu is currently a Ph.D. candidate in computer science in Nanjing University of Aeronautics and Astronautics. He received the B.S. degree in 2005 from Nanjing University of Aeronautics and Astronautics. In the same year, he was admitted to study for an M.Sc. degree in Nanjing University of Aeronautics and Astronautics without entrance examnation. His research interests focus on machine learning nd data mining.

Jian-Dong Wang graduated from Radio Department of Shanghai Jiaotong University in 1967. He is now a professor and Ph.D. supervisor of the College of Information Science and Technology, NUAA. His research interests include machine learning, data mining and information security.

Tao Li is currently a Ph.D. candidate in in computer science in Nanjing University of Aeronautics and Astronautics. He received his B.S. degree in computer science and M.S. degree in system analysis and integration from Nanjing University of Information Science and Technology, in 1999 and 2002 respectively. His research interests include recommender system and machine learning.

Appendix

In this appendix, we will give some analysis on OCVM, which mainly includes two parts, the first one (Propositions 1 and 2) shows that OCVM can guarantee that the biases are unique and properly ordered under some situation; the second one (Proposition 3 and 4) shows the approximate convergence of the solution from the viewpoints of objective function and KKT conditions.

Lemma 1. *Let ρ^* be the margin of the optimal solution of* (6), *if the matrix $\widetilde{Q}$ is positive definite, ρ^* will be unique.*

Proof. According to the definition of convex function[21], if the matrix $\widetilde{Q}$ is positive definite, the objective function of (9) will be strictly convex, then there exists at most one optimal solution $\boldsymbol{\alpha}^*$, which means that the margin ρ^* is also unique. □

Proposition 1. *Let $\boldsymbol{b}^*$ be the thresholds of the optimal solution of* (6), *if the matrix $\widetilde{Q}$ is positive definite, $\boldsymbol{b}^*$ will be unique.*

Proof. Firstly, take any one k for consideration, we define $I_j^{low}(b) = \{i \in \{1,\ldots,l_j\}\colon \langle \boldsymbol{w}, \varphi(\boldsymbol{x}_i^j)\rangle - b > -\rho\}$ and $I_j^{up}(b) = \{i \in \{1,\ldots,l_j\}\colon \langle \boldsymbol{w}, \varphi(\boldsymbol{x}_i^j)\rangle - b < \rho\}$. It is easy to see that $\boldsymbol{b}_k^*$ is optimal iff it minimizes the function:

$$\boldsymbol{e}_k(b) = \sum_{j=in(k,s)}^{k} \sum_{i\in I_j^{\rm low}(b)} (\langle \boldsymbol{w}, \varphi(\boldsymbol{x}_i^j)\rangle - b + \rho)^2 + b^2 + \sum_{j=k+1}^{su(k,s,q)} \sum_{i\in I_j^{up}(b)} (-\langle \boldsymbol{w}, \varphi(\boldsymbol{x}_i^j)\rangle + b + \rho)^2. \quad \text{(A1)}$$

According to the strict convexity[21] of (A1) and the unique of the margin ρ by Lemma 1, we have that $\boldsymbol{b}_k^*$ is unique. Then we can conclude that $\boldsymbol{b}^*$ is unique. □

Proposition 2. *If $s = q-1$, the bias of the optimal solution for the primal problem* (6) *will be ordered as $\boldsymbol{b}_1^* \leqslant \boldsymbol{b}_2^* \leqslant \cdots \leqslant \boldsymbol{b}_{q-1}^*$.*

Proof. Firstly, we define the function $\boldsymbol{e}_k(b)$ as follows:

$$\boldsymbol{e}_k(b) = b^2 + \sum_{j=in(k,s)}^{k} \sum_{i\in I_j^{\rm low}(b)} (\langle \boldsymbol{w}, \varphi(\boldsymbol{x}_i^j)\rangle - b + \rho)^2 + \sum_{j=k+1}^{su(k,s,q)} \sum_{i\in I_j^{\rm up}(b)} (-\langle \boldsymbol{w}, \varphi(\boldsymbol{x}_i^j)\rangle + b + \rho)^2.$$

Then, when $s = q-1$, the derivative of $\boldsymbol{e}_k(b)$ with respect to b is

$$\boldsymbol{g}_k(b) = \frac{\partial \boldsymbol{e}_k(b)}{\partial b} = -2\sum_{j=1}^{k} \sum_{i\in I_j^{\rm low}(b)} (\langle \boldsymbol{w}, \varphi(\boldsymbol{x}_i^j)\rangle - b + \rho) + 2\sum_{j=k+1}^{q} \sum_{i\in I_j^{\rm up}(b)} (-\langle \boldsymbol{w}, \varphi(\boldsymbol{x}_i^j)\rangle + b + \rho) + 2b.$$

Take any one k with $1 \leqslant k < q-1$ for consideration, and suppose $\boldsymbol{b}_k^* > \boldsymbol{b}_{k+1}^*$. Since $\boldsymbol{b}_{k+1}^*$ is strictly to the left of the bias $\boldsymbol{b}_k^*$ that minimizes $e_k(b)$, we have $g_k(\boldsymbol{b}_{k+1}^*) < 0$. Since $\boldsymbol{b}_{k+1}^*$ is a minimizer of $e_{k+1}(b)$, we also have $g_{k+1}(\boldsymbol{b}_{k+1}^*) \geqslant 0$. Thus we have $g_{k+1}(\boldsymbol{b}_{k+1}^*) - g_k(\boldsymbol{b}_{k+1}^*) > 0$, but by the formulation of $g_k(b)$, we get

$$\begin{aligned} & g_{k+1}(\boldsymbol{b}_{k+1}^*) - g_k(\boldsymbol{b}_{k+1}^*) \\ = & -2 \sum_{i\in I_{k+1}^{\rm low}(\boldsymbol{b}_{k+1})} (\langle \boldsymbol{w}, \varphi(\boldsymbol{x}_i^{k+1})\rangle - \boldsymbol{b}_{k+1}^* + \rho) - \\ & 2 \sum_{i\in I_{k+1}^{\rm up}(\boldsymbol{b}_{k+1})} (-\langle \boldsymbol{w}, \varphi(\boldsymbol{x}_i^{k+1})\rangle + \boldsymbol{b}_{k+1}^* + \rho) \leqslant 0 \end{aligned}$$

so $\boldsymbol{b}_k^* \leqslant \boldsymbol{b}_{k+1}^*$, and similarly we can get $\boldsymbol{b}_1^* \leqslant \boldsymbol{b}_2^* \leqslant \cdots \leqslant \boldsymbol{b}_{q-1}^*$. This completes the proof. □

Proposition 3. *If Algorithm* 1 *terminates at the* τ*-th iteration, and suppose that the optimal objective for dual problem* (10) *is* p^*, *we will have that* $\max\left\{\frac{R_\tau^2}{p^*-\tilde{\kappa}}, \frac{p^*-\tilde{\kappa}}{R_\tau^2}\right\} \leqslant (1+\epsilon)^2$.

Proof. If Algorithm 1 terminates at the τ-th iteration, we will obtain an $(1+\epsilon)$-approximation solution, that is $R_\tau \leqslant R_{MEB(\tilde{S}_\tau)} \leqslant (1+\epsilon)R_\tau$, namely,

$$(R_\tau)^2 \leqslant (R_{MEB(\tilde{S}_\tau)})^2 \leqslant ((1+\epsilon)R_\tau)^2. \tag{A2}$$

Since the optimal objective for dual problem (10) is p^*, we can get $(R_{MEB(\tilde{S}_\tau)})^2 = p^* - \tilde{\kappa}$ by the relationship between (1) and (4), then take it into (A2), it is converted to $(R_\tau)^2 \leqslant p^* - \tilde{\kappa} \leqslant ((1+\epsilon)R_\tau)^2$, Hence, we have that

$$\max\left\{\frac{R_\tau^2}{p^* - \tilde{\kappa}}, \frac{p^* - \tilde{\kappa}}{R_\tau^2}\right\} \leqslant (1+\epsilon)^2.$$

□

The Proposition 3 illuminates that the solution of Algorithm 1 is $(1+\epsilon)^2$-approximation of the optimal objective.

Proposition 4. *If Algorithm 1 terminates at the* τ*-th iteration, for each training sample* $\boldsymbol{z}_\ell$ *in* $\widetilde{S}$, *we will have that*

$$\boldsymbol{t}_\ell(\langle \boldsymbol{w}_\tau, \varphi(\boldsymbol{x}_\ell)\rangle + \boldsymbol{b}_{\phi(\ell)}) - \rho_\tau \geqslant -\max\left\{\left(\epsilon + \frac{\epsilon^2}{2}\right)\tilde{\kappa}^2, \boldsymbol{\Delta}_{\ell,\ell}\right\}.$$

Proof. Firstly, according to (5), (10) and (13), we will have that

$$(R_\tau)^2 = \tilde{\kappa} - \rho_\tau. \tag{A3}$$

Suppose that $\tilde{S}_\tau^S$ is the support vector set of $\tilde{S}_\tau$, $\forall \boldsymbol{z}_\ell \in \tilde{S} \setminus \tilde{S}_\tau^S$, by (12) and (13), we will have that

$$\begin{aligned}\|\boldsymbol{c}_\tau - \boldsymbol{z}_\ell\|^2 = &\sum_{\boldsymbol{z}_{i_1}, \boldsymbol{z}_{i_2} \in \tilde{S}_\tau} \boldsymbol{\alpha}_{i_1}^\tau \boldsymbol{\alpha}_{i_2}^\tau \boldsymbol{t}_{i_1} \boldsymbol{t}_{i_2} \widetilde{K}(\boldsymbol{x}_{i_1}, \boldsymbol{x}_{i_2}) + \\ &\widetilde{K}(\boldsymbol{z}_\ell, \boldsymbol{z}_\ell) - 2\sum_{\boldsymbol{z}_i \in \tilde{S}_\tau} \boldsymbol{\alpha}_i^\tau (\boldsymbol{t}_i \boldsymbol{t}_\ell K(\boldsymbol{x}_i, \boldsymbol{x}_\ell) + \\ &\boldsymbol{T}_{i,\ell} + \boldsymbol{\Delta}_{i,\ell}) \\ = &\rho_\tau + \tilde{\kappa} - 2\boldsymbol{t}_\ell(\langle \boldsymbol{w}_\tau, \varphi(\boldsymbol{x}_\ell)\rangle + \boldsymbol{b}_{\phi(\ell)}).\end{aligned} \tag{A4}$$

Further, $\forall \boldsymbol{z}_\ell \in ((\tilde{S} \setminus \tilde{S}_\tau^S) \cap B(\boldsymbol{c}_\tau, R_\tau))$, we will hav that $\|\boldsymbol{c}_\tau - \boldsymbol{z}_\ell\|^2 \leqslant {R_\tau}^2$, then by (A2) and (A3), we wil have $\rho_\tau + \tilde{\kappa} - 2\boldsymbol{t}_\ell(\langle \boldsymbol{w}_\tau, \varphi(\boldsymbol{x}_\ell)\rangle + \boldsymbol{b}_{\phi(\ell)}) \leqslant \tilde{\kappa} - \rho_\tau$, whicl can be rewritten as

$$\boldsymbol{t}_\ell(\langle \boldsymbol{w}_\tau, \varphi(\boldsymbol{x}_\ell)\rangle + \boldsymbol{b}_{\phi(\ell)}) - \rho_\tau \geqslant 0.$$

And $\forall \boldsymbol{z}_\ell \notin B(\boldsymbol{c}_\tau, R_\tau)$, we have that

$$R_\tau^2 < \|\boldsymbol{c}_\tau - \boldsymbol{z}_\ell\|^2 \leqslant ((1+\epsilon)R_\tau)^2 \tag{A5}$$

then take (A2) and (A4) into (A5), it is converted to

$$\begin{aligned}\tilde{\kappa} - \rho_\tau &\leqslant \rho_\tau + \tilde{\kappa} - 2\boldsymbol{t}_\ell(\langle \boldsymbol{w}_\tau, \varphi(\boldsymbol{x}_\ell)\rangle + \boldsymbol{b}_{\phi(\ell)}) \\ &\leqslant (1+\epsilon)^2(\tilde{\kappa} - \rho_\tau)\end{aligned} \tag{A6}$$

and as $R_\tau^2 \leqslant \tilde{\kappa}$, (A6) can be converted further as

$$0 > \boldsymbol{t}_\ell(\langle \boldsymbol{w}_\tau, \varphi(\boldsymbol{x}_\ell)\rangle + \boldsymbol{b}_{\phi(\ell)}) - \rho_\tau \geqslant -\left(\epsilon + \frac{\epsilon^2}{2}\right)\tilde{\kappa}^2.$$

Next, $\forall \boldsymbol{z}_\ell \in \tilde{S}_\tau^S$, by (12) and (13), we will have tha

$$\begin{aligned}\|\boldsymbol{c}_\tau - \boldsymbol{z}_\ell\|^2 = &\sum_{\boldsymbol{z}_{i_1}, \boldsymbol{z}_{i_2} \in \tilde{S}_\tau} \boldsymbol{\alpha}_{i_1}^\tau \boldsymbol{\alpha}_{i_2}^\tau \boldsymbol{t}_{i_1} \boldsymbol{t}_{i_2} \widetilde{K}(\boldsymbol{x}_{i_1}, \boldsymbol{x}_{i_2}) + \\ &\widetilde{K}(\boldsymbol{z}_\ell, \boldsymbol{z}_\ell) - 2\sum_{\boldsymbol{z}_i \in \tilde{S}_\tau} \boldsymbol{\alpha}_i^\tau (\boldsymbol{t}_i \boldsymbol{t}_\ell K(\boldsymbol{x}_i, \boldsymbol{x}_\ell) + \\ &\boldsymbol{T}_{i,\ell} + \boldsymbol{\Delta}_{i,\ell}) \\ = &\rho_\tau + \tilde{\kappa} - 2\boldsymbol{t}_\ell(\langle \boldsymbol{w}_\tau, \varphi(\boldsymbol{x}_\ell)\rangle + \boldsymbol{b}_{\phi(\ell)}) - 2\boldsymbol{\alpha}_\ell \boldsymbol{\Delta}_{\ell,\ell}\end{aligned}$$

similar with the analysis of the case $\forall \boldsymbol{z}_\ell \in ((\tilde{S} \setminus \tilde{S}_\tau^S)$ $B(\boldsymbol{c}_\tau, R_\tau))$, it is easy to show that

$$\boldsymbol{t}_\ell(\langle \boldsymbol{w}_\tau, \varphi(\boldsymbol{x}_\ell)\rangle + \boldsymbol{b}_{\phi(\ell)}) - \rho_\tau \geqslant -\boldsymbol{\alpha}_\ell \boldsymbol{\Delta}_{\ell,\ell} \geqslant -\boldsymbol{\Delta}_{\ell,\ell}.$$

So summarizing the three cases, $\forall \boldsymbol{z}_\ell \in \tilde{S}$, we ca conclude that $\boldsymbol{t}_\ell(\langle \boldsymbol{w}_\tau, \varphi(\boldsymbol{x}_\ell)\rangle + \boldsymbol{b}_{\phi(\ell)}) - \rho_\tau \geqslant -\max\{(\epsilon$ $\frac{\epsilon^2}{2})\tilde{\kappa}^2, \boldsymbol{\Delta}_{\ell,\ell}\}$.

The Proposition 4 illuminates that the solution Algorithm 1 will satisfy the loose KKT conditions.

第30卷 第2期 北京理工大学学报 Vol.30 No.2
2010年2月 Transactions of Beijing Institute of Technology Feb. 2010

标准模型下基于身份的两方认证密钥协商协议

任勇军， 王建东， 庄毅， 王箭， 徐大专
(南京航空航天大学 信息科学与技术学院，江苏，南京 210016)

摘 要：采用MTI协议族的思想，设计了一个新的标准模型下基于身份的两方认证密钥协商协议IBAKE，并形式化证明了该协议的安全性．与现有的标准模型下基于身份的密钥协商协议相比，IBAKE协议在计算效率、通信效率等方面性能更加优越．

关键词：基于身份的密码学；认证密钥协商；标准模型
中图分类号：TP 309 **文献标志码**：A **文章编号**：1001-0645(2010)02-0174-05

Identity-Based Authenticated Key Agreement Protocol for Two-Party in the Standard Model

REN Yong-jun， WANG Jian-dong， ZHUANG Yi， WANG Jian， XU Da-zhuan
(College of Information Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing, Jiangsu 210016, China)

Abstract: In this paper the idea of the MTI protocols is adopted to devise a new identity-based authenticated key agreement protocol for two-party in standard model and the formal proof of its security is provided. The proposed protocol has better performances in computational and communication efficiencies compared to all known protocols under the standard model.

Key words: identity-based cryptography; authenticated key agreement; standard model

在安全通信领域，密钥协商协议具有重要的基础性作用．王圣宝等提出了第一个标准模型下基于身份的两方认证密钥协商协议[1]（记为Wang协议），但是其安全性证明是不完整的．该协议使用了一个密钥抽取函数 H_2，但没有对其性质进行说明，在证明中也没有使用该函数．实际上，密钥抽取函数兼有随机提取器的功能[2]．在标准模型下通信方协商得到的信息首先需要使用随机提取器以获得高熵的比特串，然后再进行密钥抽取操作才能够保证会话密钥是一个在密钥空间均匀分布的比特串．Chevassut等[2]已经证明不正确地选用密钥抽取函数会产生一个固定的会话密钥值，导致协议不能抵抗攻击者的攻击．Colin等人提出采用密钥封装机制构造密钥协商协议的一般方法，并提出了3个基于身份的密钥封装方案，以此为基础构造了3个两方认证密钥协商协议[3]（分别记为IBAK1，IBAK2和IBAK3），将协议的安全性规约为密钥封装方案的安全性，但是并没有对所提出的三个基于身份的密钥封装方案的安全性进行证明，无法保证所提出协议的安全性．而且IBAK1和IBAK2协议使用了Water式的Hash函数，导致系统的公钥过大，安全性证明规约松散，协议IBAK3存在密文过长，所需传输数据量过大[4]．最近Tian等人也提出了一个标准模型下可证安全的认证密钥协商协议[5]（记为Tian协议），但使用了较弱的安全模型(BR模型)进行证明．

作者采用MTI协议族的加密-解密密钥协商思想，并根据密钥抽取函数的功能，将密钥抽取阶段细化为随机提取和密钥抽取两个步骤，以Kiltz等人的选择密文安全(chosen ciphertext attack 2,

收稿日期：2008-11-06
基金项目：国家“八六三”计划项目(2006AA706103)；国家部委预研项目(C0720061360)
作者简介：任勇军(1974—)，男，博士后，讲师，E-mail：renyj100@126.com．

CCA2) 的基于身份的加密 (identity-based encryption, IBE) 方案[4] 为基础,设计了一个新的标准模型下基于身份的认证密钥协商协议 IBAKE,使用改进的 Canetti-Krawczyk 模型[6] (记为 CK2005 模型) 形式化证明了该协议的安全性.

1　形式化安全模型

Krawczyk 指出 CK2001 模型[7] 不能抵抗 KCI 攻击和提供前向安全性,改进了 CK2001 模型,记为 CK2005 模型[6],作者使用该模型对 IBAKE 协议进行形式化证明.

CK2005 模型包括了协议参与者集合 $\{P_1, P_2, \cdots, P_n\}$, 每个参与者被模拟为一组预言机,执行的一个协议实例称为一个会话. 如果某个预言机 $M^n_{I,J}$ 发出的每条消息都相继被传送到另外一个预言机 $M^n_{J,I}$, 并且其应答消息也被传回到 $M^n_{I,J}$, 作为与其会话脚本记录对应的下一条消息,就说这两个预言机之间拥有匹配会话. 模型通过一个在挑战者和攻击者 E 之间的游戏来定义密钥协商协议的安全性. 攻击者 E 被允许进行 Send, Corrupt, session-key, session-state, session-expiration 和 Test 等预言机查询,并且这些查询可以是无序和自适应的. 在 Test 查询中,预言机通过投掷一枚公平硬币 $b \in \{0,1\}$ 来回答查询:若投币结果为 0,那么它返回协商获得的会话密钥;否则,它返回会话密钥空间 $\{0,1\}^k$ 上的一个随机值. k 表示会话密钥的比特长度. 最终,攻击者 E 输出一个对 b 的判断(记为 b'). 若 $b' = b$, 那么则称攻击者 E 赢得了此游戏,其获胜优势为: $A_E = |2P[b = b'] - 1|$.

定义 1　安全密钥协商协议. 若一个密钥协商协议满足如下两个条件: ①任何两个未腐化的协议参加者如果拥有匹配会话,那么它们就能够计算获得一个相同的会话密钥; ②对于任何恶性攻击者 E, A_E 是可忽略的,那么称该协议是一个安全的密钥协商协议.

2　新协议

2.1　IBAKE 协议描述

系统内存在一个私钥生成中心 (private key generator, PKG) 负责为用户生成和安全分发长期私钥,两个用户 A 和 B 希望通过 IBAKE 协商达成一个共享会话密钥. PKG 选取阶为素数 p 的乘法交换群 G_1, G_2, 双线性对 $e: G_1 \times G_1 \to G_2$, G_1 上的生成元 f 和 g, 随机整数 $\alpha, \beta, \gamma \in Z_p$, 计算 $g_1 = g^{\alpha}$, $v_1 = e(g, g)^{\beta}$, $v_2 = e(g, g)^{\gamma}$. 用户 A 和 B 的长期私钥 $d_{\mathrm{id}_i} = (s_{i,1}, s_{i,2}, d_{i,1}, d_{i,2})$, $i \in \{A, B\}$, 其中 $s_{i,1}$, $s_{i,2} \in \mathbf{Z}_p$, $d_{i,1} = g^{\frac{\beta - s_{i,1}}{\alpha - i}}$, $d_{i,2} = g^{\frac{\gamma - s_{i,2}}{\alpha - i}}$.

IBAKE 协议由 3 个阶段组成:系统建立,私钥生成和密钥协商阶段. 其中,系统建立和私钥生成阶段与 Kiltz 基于身份的加密方案[4] 完全相同. 密钥协商阶段由 3 部分组成:加密、解密和计算.

2.1.1　加密

A 和 B 分别随机选取 r_A 和 r_B ($r_A, r_B \in Z_p$), 然后分别执行如下的加密操作.

①A 计算. $C_{A1} = (g_1 g^{-B})^{r_A}$, $C_{A2} = e(g, g)^{r_A}$, $t_A = \mathrm{TCR}(C_{A1}, C_{A2})$, $K_A' = (v_1^{t_A} v_2)^{r_A}$; 随机选取短期私钥 $y_A \in Z_p$, 计算 $Y_A = f^{y_A}$; 然后将 (A, $C_A = (C_{A1}, C_{A2})$, Y_A) 发送给 B.

②B 计算. $C_{B1} = (g_1 g^{-A})^{r_B}$, $C_{B2} = e(g, g)^{r_B}$, $t_B = \mathrm{TCR}(C_{B1}, C_{B2})$, $K_B' = (v_1^{t_B} v_2)^{r_B}$; 随机选取短期私钥 $y_B \in Z_p$, 计算 $Y_B = f^{y_B}$; 然后将 (B, $C_B = (C_{B1}, C_{B2})$, Y_B) 发送给 A.

2.1.2　解密

A 和 B 接收到消息后,分别使用自己的私钥执行解密操作.

①A 接收到 C_B 后计算. $t_B = \mathrm{TCR}(C_{B1}, C_{B2})$, $K_B' = e(C_{B1}, d_{A,1}^{t_B} d_{A,2}) C_{B2}^{t_B s_{A,1} + s_{A,2}}$.

②B 接收到 C_A 后计算. $t_A = \mathrm{TCR}(C_{A1}, C_{A2})$, $K_A' = e(C_{A1}, d_{B,1}^{t_A} d_{B,2}) C_{A2}^{t_A s_{B,1} + s_{B,2}}$.

2.1.3　计算

A 和 B 分别得到 K_B' 和 K_A' 后进行如下计算.

① A 随机提取: $K_A'' = \mathrm{Exct}_k(K_A')$, $K_B'' = \mathrm{Exct}_k(K_B')$, $K_{AB}'' = \mathrm{Exct}_k(Y_B^{y_A})$.

A 密钥抽取: $s = A \| C_A \| Y_A \| B \| C_B \| Y_B$, $K_A = \mathrm{Expd}_{K_A''}(s) \oplus \mathrm{Expd}_{K_B''}(s) \oplus \mathrm{Expd}_{K_{AB}''}(s)$, 然后删除除 s 和 K_A 之外的其它临时信息.

② B 随机提取: $K_B'' = \mathrm{Exct}_k(K_B')$, $K_A'' = \mathrm{Exct}_k(K_A')$, $K_{BA}'' = \mathrm{Exct}_k(Y_A^{y_B})$.

B 密钥抽取: $s = A \| C_A \| Y_A \| B \| C_B \| Y_B$, $K_B = \mathrm{Expd}_{K_B''}(s) \oplus \mathrm{Expd}_{K_A''}(s) \oplus \mathrm{Expd}_{K_{AB}''}(s)$, 然后删除除 s 和 K_B 之外的其它临时信息.

其中 TCR 是哈希函数 (target collision resistant), $\mathrm{Exct}_k(\cdot): K \to U_1$ 是一个 (m, ε) - 随机提取器, $\{\mathrm{Exdp}_k(\cdot)\}_{K \in U_1}: \{0,1\}^{\sigma} \to U_2$ 是一个伪随机函数族[2]. 需要注意的是在 session-state 查询中,预