

文章编号: 1001-0920(2015)07-1315-06

DOI: 10.13195/j.kzyjc.2014.0757

一种基于信念修正思想的SVR增量学习算法

陈海燕, 丰文安, 王建东, 王寅同, 孙博

(南京航空航天大学 计算机科学与技术学院, 南京 210016)

摘要: 针对实际应用中数据的批量到达, 以及系统的存储压力和学习效率低等问题, 提出一种基于信念修正思想的SVR增量学习算法. 首先从历史样本信息中提取信念集, 根据信念集和新增数据的特点选择相应的信念集建立支持向量回归模型并进行预测; 然后对信念集进行修正, 调整当前认知状态, 使该算法对在线和批处理增量学习都有很好的适应性. 在标准数据集上的测试验证了算法的良好性能; 在某机场噪声实测数据上的对比实验也表明, 该算法的性能明显优于传统学习算法和一般增量学习算法.

关键词: 信念修正; 增量学习; 支持向量回归; 认知状态; 机场噪声

中图分类号: TP301.6

文献标志码: A

An SVR-based incremental learning algorithm based on belief revision

CHEN Hai-yan, FENG Wen-an, WANG Jian-dong, WANG Yin-tong, SUN Bo

(School of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China. Correspondent: FENG Wen-an, E-mail: fengwenan@126.com)

Abstract: In order to reduce the storage cost and improve the learning efficiency when data arrives in batches in practical applications, an SVR-based incremental learning algorithm based on belief revision is proposed. In this algorithm, belief sets are extracted from the historical samples firstly, and appropriate belief sets are selected to establish the prediction SVR model according to the characteristics of belief sets and the new incremental data set. Then, the belief sets are revised and the current epistemic state is adjusted to adapt the new environment. The algorithm perform well when applied to online learning and incremental learning. Experiments conducted on the standard data sets and the airport noise data set demonstrate that the proposed algorithm is effective and superior to the traditional methods and the general incremental learning algorithms.

Keywords: belief revision; incremental learning; SVR; epistemic state; airport noise

0 引言

增量学习是机器学习领域的一个研究热点, 广泛应用于众多领域. 随着信息技术的发展, 数据量越来越庞大, 系统无法一次分析所有数据, 或者由于数据分批次到达, 系统无法获得一个完备的数据集, 从而要求模型应具有良好的增量学习功能, 能够快速有效地捕捉新知识, 不断巩固现有的学习成果, 提高模型的预测能力.

近年来, 随着增量学习算法研究的不断深入, 很多新颖的改进方法相继涌现出来. 有的方法着力于提升增量学习效率, 如 Syed 等^[1]提出将新增样本与当前 SVM 模型的支持向量组成新的训练集, 缩短了模型更新的时间; Mitra 等^[2]提出了一种数据约减的方法,

选取分类边界附近的样本进行训练, 减少了训练量, 降低了内存需求; Okamoto 等^[3]利用优化后的梯度下降法选取新增样本和历史样本进行训练, 在保证增量学习能力的前提下缓解了存储压力, 并提高了运行效率; Zhu 等^[4]优化了 ISVM 的矩阵运算, 提高了增量学习的效率; Zheng 等^[5]提出了一种在线 SVM 增量学习算法, 能够快速有效地处理大数据问题; Du 等^[6]提出了一种快速增量学习算法, 基于聚类的思想对新增样本进行快速有效的处理. 也有的方法旨在提高模型预测能力, 如 Schlimmer 等^[7]提出了特征加权以及在增量学习过程中抓取实例特征并标记, 增强模型的鲁棒性; Zhang 等^[8]提出使用样本点与两类中心点的距离的比值来约减数据样本; Cauwenberghs^[9]提出了一种

收稿日期: 2014-05-15; **修回日期:** 2014-09-09.

基金项目: 国家863计划项目(2012AA063301); 国家自然科学基金重点项目(61139002); 中央高校基本科研业务费专项资金项目(NS2012134, NZ2013306); 江苏省博士后计划项目(1301013A); 中国民航信息技术科研基地开放基金项目(CAAC-ITRB-201203).

作者简介: 陈海燕(1979—), 女, 讲师, 博士, 从事数据挖掘与民航信息化的研究; 王建东(1945—), 男, 教授, 博士生导师, 从事数据挖掘、机器学习和信息安全等研究.

增量学习精确算法,通过新增数据不断地对原有的模型参数进行修正;Li等^[10]引入多核学习方法和惩罚因子来解决大数据学习问题,提高了增量SVM性能;Elwell等^[11]提出了一种非平稳环境下的增量学习算法,增强了模型的预测能力.

本文从信念修正角度分析并提出一种新的增量学习算法,抛弃传统的模型依赖的增量学习算法,根据信息关联度从历史样本信息中提取信念集,进而构造学习模型,根据信念特征对新增样本信息归类并对信念集进行修正;同时,提出了相似情形的概念,强调周边样本在预测中起到的重要性.基于信念修正的增量学习算法可以在保证并提高算法精度的同时,大大缩短所需的时间.

1 信念修正

1.1 信念修正的概念

信念修正起源于哲学和逻辑学,后来渐渐发展成为人工智能领域的研究热点,并延伸为数据库和数据挖掘的研究方法.信念修正是变更信念来采纳新的信息片段的过程,人的知识或信念不是一成不变的,而是随着新的认知信息不断调整自己的认知状态以达到一种理想化的平衡状态.信念修正的研究成果中,较为重要的有真值维护方法、序列极限方法、迭代方法和AGM(Alchourron, Gardenfors 和 Makinson)方法.其中AGM理论在计算机科学和人工智能领域有很大影响,使用集合论来刻画认知状态、认知输入信息以及3种信念修正函数,便于进行技术处理.

信念修正一般遵循3个原则:一致性原则,新消息优先原则和最小改变原则.一致性原则是指认知状态是一种平衡状态,必须保证信念集的一致性以及逻辑闭包;新消息优先原则强调新信息的重要性,需要改变旧信念来适应新信息;在信念变更或修正过程中,需要遵循最小改变原则,以最小的代价接受新信息,前后的知识应尽可能地类似^[12].

1.2 信念修正的变化形式

信念变化包括3个基本形式:扩充、收缩和修正.

令 Γ 表示当前认知状态,称为信念集合; α 表示任意的语句.

扩充:将一个语句 α 及其增加的逻辑结论加入信念集合 Γ 得到新的集合 Γ_{α}^{+} .

收缩:将 Γ 中某个语句 α 及其相关的逻辑结论删除,得到的集合 Γ_{α}^{-} 仍是一个逻辑闭包.

修正:将一个与 Γ 不一致的语句 α 加入 Γ 中,需要将 Γ 中原有的某些语句删除,使得修改后的信念集 Γ_{α}^{*} 与 α 是协调的.

这3个基本形式只有扩充可以确切定义;对于收缩和修正,AGM理论给出了若干假设.根据Harper等

式,可以将收缩函数转化为修正函数 $\Gamma_{\alpha}^{-} = \Gamma \cap \Gamma_{\neg\alpha}^{*}$,这里不再赘述.下面重点考虑修正过程假设的部分基本条件^[13]:

- 1) 对于任意的 α 和信念集 Γ ,根据AGM理论, $\Gamma_{\alpha}^{+} = \{\beta : (\Gamma \cup \{\alpha\}) \vdash \beta\}$, Γ_{α}^{*} 是信念集合;
- 2) $\alpha \in \Gamma_{\alpha}^{*}$,在信念修正过程中,新信息 α 不可丢失;
- 3) 当 $\neg\alpha \notin \Gamma$, $\Gamma_{\alpha}^{*} = \Gamma_{\alpha}^{+}$,如果 α 与 Γ 保持一致性,则修正等价于扩充;
- 4) 当 $\neg\alpha \in \Gamma$, $\Gamma_{\alpha}^{*} \subseteq \Gamma_{\alpha}^{+}$,如果 α 与 Γ 矛盾,则需要对 Γ 中原有的某些语句进行约减;
- 5) Γ_{\perp} 表示所有语句组成的集合, $\Gamma_{\alpha}^{*} = \Gamma_{\perp}$,当且仅当 $\vdash \neg\alpha$,除非 α 常假,否则 Γ_{α}^{*} 需要保持一致性;
- 6) 如果 $\vdash \neg\alpha \leftrightarrow \beta$,则 $\Gamma_{\alpha}^{*} = \Gamma_{\beta}^{*}$.

2 基于信念修正思想的SVR增量学习算法

2.1 数据的一致性原则

数据集可以视为向量的广义集合,由于数据集中可能会出现重复数据,在增量学习中,这些重复数据的价值相对于其他数据而言会贬值,在同等训练规模的条件下,需要保证数据质量,则必须对这些冗余样本进行约减.本文提到的所有集合都是经过重复数据约减操作后得到的数据集,在下文表述中不再特别说明.

本文研究的是一种基于信念修正思想的SVR增量学习算法,而在SVR建模过程中,最大化间隔、建立超平面以及典型核函数的构造形式,都与数据间的欧氏距离有关,所以相似样本和不一致样本集是根据样本间的欧氏距离来定义的^[14].

假设 s^p 是样本 s 的属性向量, s^t 是样本 s 的目标值,样本 s_i 与 s_j 之间的关联可以表如下:

属性关联

$$d_{s_i s_j} = \sqrt{|s_i^p - s_j^p|^2}; \quad (1)$$

目标关联

$$d'_{s_i s_j} = \sqrt{|s_i^t - s_j^t|^2}. \quad (2)$$

假设存在数据集 E 、 S 以及一条数据 s .其中: $E \neq \emptyset$, $S \subseteq E$, $s \in S$, d_{\min} 与 d'_{\min} 是根据不同数据集和不同问题设置的阈值,通过观察各数据集中数据分布的特点选择相应阈值区间,然后交叉验证分别得到合适的阈值.于是 s 在 S 中的相似样本子集记作 $s_{\tilde{S}}$,定义为

$$s_{\tilde{S}} = \{s_i | (s_i \in S) \wedge (d_{ss_i} \leq d_{\min})\}; \quad (3)$$

s 在 S 中的最相似样本子集记作 $s_{\tilde{S}^*}$,定义为

$$s_{\tilde{S}^*} = \{s_i | (s_i \in S) \wedge \forall s_j \in S (d_{ss_i} \leq d_{ss_j})\}; \quad (4)$$

函数 $f_S^B(s)$ 表示 B 中样本在 S 中的最相似样本集包含 s 的集合,定义为

$$f_S^B(s) = \{b_i | (s \in (b_i)_{\tilde{S}}) \wedge (b_i \in B)\}; \quad (5)$$

s 的不一致样本集记作 $\neg s$, 定义为

$$\neg s = \{s_i | (s_i \in S_E^{\sim}) \wedge (d'_{ss_i} > d'_{\min})\}; \quad (6)$$

S 的不一致样本集记作 $\neg S$, 定义为

$$\neg S = \{s_i | (s_i \in S_E^{\sim}) \wedge (d'_{ss_i} > d'_{\min}) \wedge (s \in S)\}. \quad (7)$$

不一致样本其实并非是相互矛盾的,也可能是当前认知状态所不能解读的.例如,数据集的属性不是完备的,即除了这些属性之外,还会有其他的属性影响着目标属性,而这些没有被归纳进来的属性可能是导致这种不一致性的根源.

2.2 数据集上的信念集构造

信念集含有很多信念,这些信念其实能分成很多种类.例如,人有衣食住行各方面的知识,在大脑中已自行划分了区域,这样便能有针对性地调用相应知识解决问题.而且知识的频繁更新需要人对新知识具有良好的适应能力,包括处理能力和处理速度,最终目的是快速有效地提高人的认知水平^[15].

如果将数据看成信息,从信息中凝聚出知识,则可以将模型看成系统的认知状态.知识应尽量覆盖各个方面以应对各种问题,而解决问题时只需要运用某一个或几个区域的知识.若过多运用其他知识,则可能会起到相反的效果.同样,对于模型而言,应使训练数据分布在整个数据集上,能大体反映整个数据集的规律,以达到较高的总体预测精度;而对于单条数据的预测,这样选取训练数据并不一定能达到好的效果^[16-17].

对于一个数据集而言,当前所有属性的信息并不是完备的;当某些属性值的变动触发某种事件时,可能会由量变产生质变,从而使其遵循另外一种规律,这种规律的变化不易察觉.因此,对于单条数据样本,真正影响其预测精度的是其附近的样本,较远样本的影响并不明显,甚至会起到负面影响.直观上看,训练属性值较为接近的训练样本能更好地反映测试样本的真实情况,而差异较大的样本则可能是遵循另外一种规律,如果一起训练,则会影响最终的预测效果.

在增量学习中,模型表示系统当前的认知状态,而模型的性能体现了当前的认知水平,模型的更新速度则体现了系统的学习效率.根据信念修正思想,将现有信息分为 k 个簇,对各簇信息进行归纳,建立相应的模型,这里的信息簇为信念集,模型为认知状态.这样便可根据预测样本的特点选择相应的模型以进行更有效的预测.

定义 ird 为信息关联度, N 为信息量. k 应随着 ird 的增大而单调递减.当 ird = 0 时, $k = N$, 即信息之间没有关联,所以每条数据是一个簇;当 ird = 1 时, $k = 1$, 即信息之间都有关联,所以全部数据视为一个

簇.

不同的数据集其 ird 也不同,本文在数据集上对每个数据进行最近邻样本采集,计算其与最近邻样本的欧氏距离 d_i , 有

$$n_i = \begin{cases} 1, & d_i \leq d_{\min}; \\ 0, & d_i > d_{\min}. \end{cases}$$

统计拥有相似样本的样本个数 $n = \sum_{i=1}^N n_i$, 数据集的信息关联度为 ird = n/N .

定义数据集 $R = \{r_i | i = 1, 2, \dots, N\}$, 为便于处理,定义任意两个信念集之间的交集为空,即

$$R^* = \left\{ R_i \mid \left(R = \bigcup_{i=1}^k R_i \right) \wedge (\forall x, y, 1 \leq x \leq y \leq k, R_x \cap R_y = \emptyset) \right\}.$$

定义操作 $D_R(R^+)$: 假设 R^+ 是集合的集合, $\forall R^a, R^b \in R^+, R^a \cap R^b \neq \emptyset$, 如果 $|R^a| \leq |R^b|$, 则 $R^a = R^a - R^b$.

下面开始构造信念集.

Step 1: $x = 0, r^0 = \{(r_i)_{\tilde{R}} | r_i \in R\}$, 将数据集中的每个样本及其相似样本抽取出来.

Step 2: $\exists \alpha, \beta \in r^x \wedge |\alpha \cap \beta| \geq p * (|\alpha| + |\beta|)$, 有 $r^{x+1} = (r^x - \{\alpha\} - \{\beta\}) \cup \{\alpha \cup \beta\}$. 根据 ird 调整 p 的值, 将相似样本集按要求进行合并, 如果 $r^{x+1} = r^x$, 则令 $r = r^x$, 执行 **Step 3**; 否则令 $x = x + 1$, 继续执行 **Step 2**.

Step 3: $r_c = \{\bar{r}_i | r_i \in r\}$, k 为整数, 且 $1 \leq k \leq |r_c|$, $\exists r_c^0 (r_c \subseteq r_c \wedge |r_c^0| = k)$, 选取模较大的 k 个相似样本集, 并求解样本集中心.

Step 4: 假设各相似样本集中心 $r_c^0 = \{r_i^0 | i = 1, 2, \dots, k\}$, $y = 0, R^0 = \{R_i^0 | (R_i^0 = f_{r_c^0}^R(r_i^0)) \wedge (r_i^0 \in r_c^0)\}$; 消除 R^0 中重复样本 $D_R(R^0)$.

Step 5: $r_c^{y+1} = \{\bar{R}_i^y | i = 1, 2, \dots, k\}$, 分别求解新的相似样本集中心; $R^{y+1} = \{R_i^{y+1} | (R_i^{y+1} = f_{r_c^{y+1}}^R(\bar{R}_i^y)) \wedge (\bar{R}_i^y \in r_c^{y+1})\}$; 消除 R^{y+1} 中重复样本

$D_R(R^{y+1})$; $dt^{y+1} = \sum_{i=1}^N d_{r_i r'_i}, r'_i \in (r_i)_{\tilde{r}_c^{y+1}}$, 根据与样本集中心的相似性将数据集中的数据划分成 k 个簇, 如果 $(|dt^{y+1} - dt^y| \leq \xi) \vee (y > t)$, 则令 $R^* = R^{y+1}$, 信念集构造完成; 否则令 $y = y + 1$, 重复执行 **Step 5**, 迭代继续对数据集重新进行划分, 直至达到规定迭代次数的上限 t 或者该标准测度函数收敛至 ξ 为止^[18].

由此可见, 本文提出的信念集构造方法能够更好地对数据集进行划分, 将相似情形样本作为重点归纳出来, 但是在构造信念集时, 并没有考虑信念集的一致性原则. 由于无法区分不一致数据的“正确性”, 本文研究的是增量学习算法, 可以随着新样本的到来而

不断对其进行修正。

2.3 基于信念修正思想的样本约减方法

数据集 D_s 可以看作当前认知的复合命题信息集合, 定义操作 D_n , 对于任意一个数据子集 $D \subseteq D_s$, 都有 $D_n(D) = \{d | D \vdash d\}$. 由于知识库的不完整性和不一致性, 可将信念集限制在有限集上, 并且对逻辑闭包性不作要求, 这时可以假设 $D_n(D) = D$; 而数据集可以认为是小型的知识库, 能够支持决策, 该假设同样成立, 只是会表现出更明显的不完整性, 具体表现为冗余、包含、循环以及冲突现象的不一致性在信息缺乏情况下可能直接导致错误的决策^[19].

对知识的处理, 应尽量保证其完整性, 而系统所能接受的知识都是有上限的. 在机器学习中, 应注重知识的应用, 发挥关键信息的作用, 摒弃无用以及错误的信息; 而当信息出现不一致时, 常常会导致系统的错误决策, 所以应减免信息不一致性的影响. 因此, 当接收到新信息时, 如果发现它与历史信息不一致, 则应根据信念修正中新消息优先原则, 对历史信息进行优化, 约减部分相关历史信息, 使得新信息能够被信念集接受; 如果没有出现不一致现象, 则将新增数据放入历史样本集, 当样本已达到存储上限时, 需要对历史不一致信息和最相似样本进行约减, 并对新信息进行优化.

假设当前数据集 $M = \{m_i | i = 1, 2, \dots, N\}$, N 是数据量, 增量样本集为 D , 存储上限为 C , 定义 $M_D^+ = M \cup D$, 增量学习之后的数据集为 M_D^* , 则有:

1) 当 $|M \cup D| \leq C$ 时

$$M_D^* = M_D^+ - (-D \cap M).$$

当 $-D \cap M = \emptyset$ 时, $M_D^* = M_D^+$, 增量过程表现为知识扩充; 当 $-D \cap M \neq \emptyset$ 时, 增量过程表现为修正过程.

2) 当 $|M \cup D| > C$ 时

$$M_D^* = M_D^+ - ((-D' \cap M) \cup d_M^{\approx}),$$

其中

$$D' = \bigcup_{i=1}^N (f_M^D(m_i)) \approx f_M^D(m_i).$$

当增量样本为 $D = \{d\}$ 时, 可对增量过程进行如下化简:

1) 当 $|M \cup d| \leq C$ 时

$$M_d^* = M_d^+ - (-d \cap M);$$

2) 当 $|M \cup d| > C$ 时

$$M_d^* = M_d^+ - ((-d \cap M) \cup d_M^{\approx}),$$

从而避免了冗余样本和不一致样本.

上述样本约减方法可以逐步削弱并消除初始信念集中样本的不一致性、新样本中的不一致性以及新

样本和旧样本的不一致性. 而且, 在保证新消息优先的同时, 根据最小改变原则, 首先, 只对部分信息进行变更, 尽可能多地保留历史知识; 然后, 样本筛选不依赖于模型. 这样, 新信息的到来将不会导致模型的大幅变化. 当系统停止运行或不进行预测工作时, 可以避免模型的更新损耗, 而样本依然可以优化. 通过这样的样本约减可以控制样本集的大小, 优化历史样本, 保证样本质量, 从而提高模型的预测精度.

2.4 一种基于信念修正思想的SVR增量学习算法(BISVR)

BISVR 算法流程如图 1 所示.

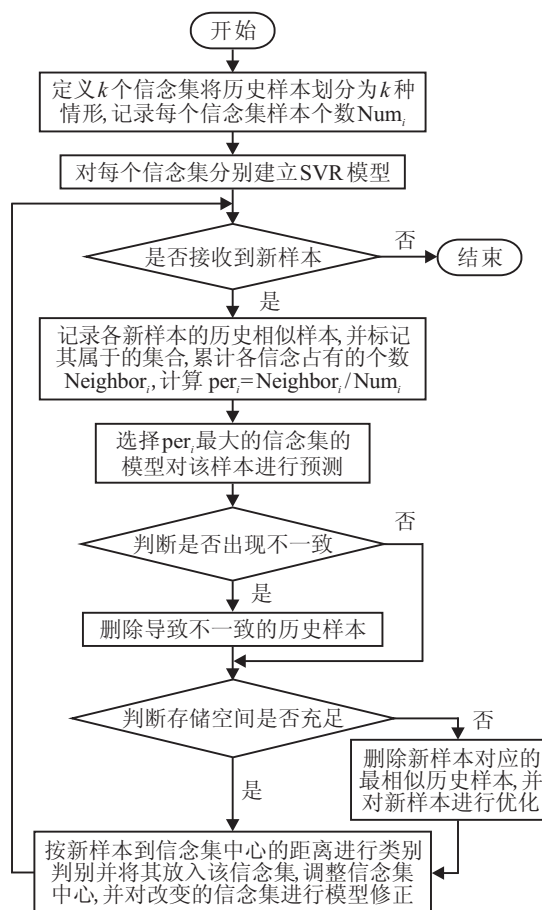


图 1 基于信念修正思想的SVR增量学习算法流程

算法步骤如下.

Step 1: 对初始样本集进行信念集构造, 并分别建立SVR模型.

Step 2: 判断是否有新增样本, 如果没有新增样本, 则算法结束; 如果有, 则记录新样本的相似样本, 根据新样本的相似样本对新样本进行信念集的判断, 并选用相应的模型进行预测.

Step 3: 根据样本约减公式删除与新样本不一致的历史样本, 并判断是否需要对新样本进行优化以及历史样本的再次约减.

Step 4: 调整信念集, 并对有变动的信念集进行模型修正.

3 实验及结果分析

3.1 算法性能验证

本文分别在 10 个标准数据集上进行实验, 如果数据集中数据量大于 5 000, 则随机抽取 5 000 条数据的数据段. 保持原始数据的有序性, 模拟数据的动态变化, 从而验证增量学习算法应对动态变化的数据的适应能力. 由于在实际应用中系统性能以及存储能力的限制, 应当限制训练集的大小.

实验 1 验证相似样本的预测能力. 为便于实验, 选取测试样本一定比例的近邻样本作为相似样本, 采用 THEIL 作为精度评价标准, 选取前 90% 作为训练样本集, 剩余的作为测试样本. Rand_SVR 算法是指分别采用将整个训练样本集作为训练集和随机抽取其中 10% 样本两种抽样方法进行训练. KNN_SVR 算法是从训练样本集中抽取 10% 测试样本的近邻样本作为训练集进行预测. 实验结果如表 1 所示.

从实验结果可以看出: KNN_SVR 算法只用 10%

表 1 相似样本预测能力验证实验结果比较 %

Database	Rand_SVR		KNN_SVR
	100	10	10
abalone	0.097 88	0.144 7	0.096 59
cloud	0.009 862	0.021 55	0.007 893
cpusmall	0.018 52	0.023 78	0.017 54
forestfire	0.800 1	1.000 0	0.757 1
housing	0.156 7	0.168 1	0.114 8
parkinsons_u	0.219 6	0.242 7	0.211 2
pyrim	0.042 45	0.074 08	0.043 20
space_ga	0.095 74	0.130 5	0.086 67
triazines	0.129 0	0.144 0	0.133 2
whitewine	0.050 75	0.073 53	0.058 86

的训练集就能够保证预测精度, 甚至超过将所有样本作为训练集的预测精度; 而随机抽取同等样本数的训练集根本无法达到这种效果. 可见, 相似样本在预测中起到了关键作用, 而其他样本则很难发挥作用, 甚至会干扰问题的处理. 本文提出的 BISVR 算法就是在 KNN_SVR 算法的基础上展开的.

实验 2 验证本文提出的 BISVR 算法的性能并与传统学习算法 TSVR、传统增量学习算法 TISVR 以及基于聚类的局部支持向量机增量学习算法 CISVR 进行对比. 对于 TISVR、CISVR、BISVR, 选取数据集中前 3/5 的数据作为初始样本集, 接下来的 1/5 作为增量样本集, 剩余的 1/5 作为测试样本集; 对于 TSVR 算法, 选取数据集中前 4/5 的数据作为训练集, 对其余 1/5 的样本进行预测; 对于 BISVR 算法, 设定存储空间大小等于初始样本集的大小, 而其他 3 种算法不做限制. 采用 THEIL 和 MSE 作为精度评价标准, 实验结果如表 2 和表 3 所示.

表 2 标准数据集实验效率比较

Database	TSVR	TISVR	CISVR	BISVR
abalone	223.86	179.90	52.35	45.96
cloud	21.29	8.98	2.91	1.57
cpusmall	662.00	400.83	202.04	146.38
forestfire	10.36	2.50	0.82	0.73
housing	13.21	9.50	2.46	2.33
parkinsons_u	1 298.83	911.46	185.28	188.20
pyrim	0.12	0.10	0.09	0.04
space_ga	218.62	139.27	82.79	79.57
triazines	0.77	0.65	0.19	0.14
whitewine	580.28	377.5	232.39	128.99

表 3 标准数据集实验精度比较

Database	THEIL				MSE			
	TSVR	TISVR	CISVR	BISVR	TSVR	TISVR	CISVR	BISVR
abalone	0.092 50	0.092 70	0.092 59	0.092 56	3.322	3.332	3.329	3.324
cloud	0.006 300	0.006 100	0.007 691	0.005 860	9.033	8.602	13.49	7.827
cpusmall	0.018 20	0.018 70	0.020 50	0.017 79	9.751	10.37	12.54	9.453
forestfire	0.863 5	0.861 1	0.081 18	0.836 0	6 497	6 490	6 453	6 385
housing	0.192 8	0.183 4	0.176 1	0.156 3	56.38	49.74	36.50	28.44
parkinsons_u	0.219 3	0.144 6	0.160 1	0.140 6	0.013 10	0.006 600	0.007 367	0.006 040
pyrim	0.042 80	0.043 00	0.059 47	0.030 52	0.003 200	0.003 200	0.006 183	0.001 638
space_ga	0.181 6	0.178 9	0.185 7	0.152 6	0.049 40	0.047 80	0.051 06	0.034 46
triazines	0.126 9	0.110 0	0.138 0	0.114 3	0.028 80	0.023 10	0.033 12	0.023 14
whitewine	0.055 50	0.055 76	0.052 99	0.053 61	0.440 5	0.401 4	0.397 5	0.405 3

从表 2 和表 3 可知: BISVR 比 TSVR 和 TISVR 在精度上有一定的优势, 尤其是在数据分布呈现一定的变化规律的数据集上优势更为明显, 如 cloud、housing、pyrim、space_ga 等数据集; CISVR 的预测能力不稳定, 而且在多数数据集上不如 BISVR 的预测精度高; BISVR 和 CISVR 对数据的处理速度都非常快, 然而 BISVR 数据的更新过程脱离模型单独进行, 在非预测阶段, 可以避免模型的及时更新, 效率会

比实验结果更快, 更符合增量学习对运行效率的要求; 而且 BISVR 能够在规模受限的样本集上尽可能地提高样本质量, 更适用于嵌入式开发. 综上所述, BISVR 在保证精度并有所提高的同时, 大大降低了时间开销, 而且预测能力比较稳定.

3.2 在机场噪声预测中的应用

本文在机场实测噪声数据上进行实验. 以某国际

机场的飞行历史记录为原始数据,数据集包含了监测点属性、飞过监测点上方航班的相关属性以及气象属性,如:点坐标、飞行类型(起、降、盘旋)、飞机型号、发动机推力、航速、飞行高度、斜距、早中晚各时段的飞行次数、温度湿度压强.因实测数据是不断到达的,使得数据分布出现层次性和阶段性,故预测模型的学习也是随之增量式开展的.

选取一段 5000 条数据,对于 TISVR、CISVR、BISVR,选取前 3/5 的数据作为初始样本集,接下来的 1/5 作为增量样本集,剩余的 1/5 作为测试样本集;对于 TSVR 算法,选取数据集中前 4/5 的数据作为训练集,对其余 1/5 的样本进行预测.

实验结果如表 4 所示.可以看出: BISVR 算法与 TSVR 和 TISVR 相比,在保证精度并有所提高的同时,大大加快了运行效率;在时间开销和预测精度上 BISVR 比 CISVR 好.

表 4 机场实测噪声数据结果比较

Algorithm	THEIL	MSE	TIME
TSVR	0.04930	55.99	1762.44
TISVR	0.05000	57.84	1147.21
CISVR	0.04942	57.43	319.06
BISVR	0.04196	41.76	271.04

4 结 论

本文提出的基于信念修正思想的 SVR 增量学习算法采用了新的样本约减方法,从数据角度利用原始样本对新样本进行优化,同时利用新样本对原始样本进行约减,保证了训练样本的质量和多样性,避免了训练样本的不一致性,并且将样本进行细化,将相似情形的样本大致归纳为同一信念集,利用多个信念集进行信念融合;在进行信念修正时,只需要对部分信念集进行调整,从而加快了知识的更新速度.该算法不仅适合批处理增量学习,也适合在线增量学习,使用这一算法分别对标准数据集数据和国内某机场的实测数据进行预测,从常用的两种精度评价标准和时间开销上可以看出,本文提出的算法能够在保证精度并有所提高的同时,大大降低了时间开销.

参考文献(References)

[1] Syed N A, Liu H, Sung K K. Incremental learning with support vector machines[C]. Proc of Int Joint Conf on Artificial Intelligence. San Jose, 1999: 641-642.
 [2] Mitra P, Murthy C A, Pal S K. Data condensation in large databases by incremental learning with support vector machines[C]. Proc of the 15th Int Conf on Pattern Recognition. Los Alamitos, 2000: 708-711.
 [3] Okamoto K, Ozawa S, Abe S. A fast incremental learning algorithm of RBF networks with long-term memory[C]. Proc of the Int Joint Conf on Neural Networks. Portland, 2003: 102-107.

[4] Zhu Z F, Zhu X Q, Guo Y F. Inverse matrix-free incremental proximal support vector machine[J]. Decision Support Systems, 2012, 53(3): 395-405.
 [5] Zheng J, Shen F R, Fan H J. An online incremental learning support vector machine for large-scale data[J]. Neural Computing and Applications, 2013, 22(5): 1023-1035.
 [6] Du H L, Teng S H, Zhu Q F. Fast SVM incremental learning based on clustering algorithm[C]. Intelligent Computing and Intelligent Systems. Piscataway, 2009: 13-17.
 [7] Schlimmer J C, Granger Jr R H. Incremental learning from noisy data[J]. Machine Learning, 1986, 1(3): 317-354.
 [8] Zhang L, Zhou W D, Jiao L C. Pre-extracting support vectors for support vector machine[C]. Proc of the 5th Int Conf on Signal Processing. Beijing: 2000: 1432-1435.
 [9] Cauwenberghs G, Poggio T. Incremental and decremental support vector machine learning[J]. Advances in Neural Information Processing Systems, 2001, 44(13): 409-415.
 [10] Li Z Y, Zhang J F, Hu S S. Incremental support vector machine algorithm based on multi-kernel learning[J]. J of Systems Engineering and Electronics, 2011, 22(4): 702-706.
 [11] Elwell R, Polikar R. Incremental learning of concept drift in nonstationary environments[J]. Neural Networks, 2011, 22(10): 1517-1531.
 [12] Gardenfors, Peter. Belief revision[M]. London: Cambridge University Press, 2003: 1-9.
 [13] Alchourron C E, Gardenfors P, Makinson D. On the logic of theory change: Partial meet contraction and revision functions[J]. J of Symbolic Logic, 1985, 50(2): 510-530.
 [14] Vapnik V. The nature of statistical learning theory[M]. New York: Springer-Verlag, 1995: 5-13.
 [15] Maynard-Reid I I P, Shoham Y. Belief fusion: Aggregating pedigreed belief states[J]. J of Logic, Language and Information, 2001, 10(2): 183-209.
 [16] Kfir-Dahav N E, Tennenholtz M. Multi-agent belief revision[C]. Proc of the 6th Conf on Theoretical Aspects of Rationality and Knowledge. San Francisco, 1996: 175-194.
 [17] Satoh K, Yamamoto K. Speculative computation with multi-agent belief revision[C]. Proc of the 1st Int Joint Conf on Autonomous Agents and Multiagent Systems. Bologna, 2002: 897-904.
 [18] Wagstaff K, Cardie C, Rogers S. Constrained k -means clustering with background knowledge[C]. Proc of the 18th Int Conf on Machine Learning. Washington DC, 2001: 577-584.
 [19] Dubois D, Lang J, Prade H. Automated reasoning using possibilistic logic: Semantics, belief revision, and variable certainty weights[J]. Knowledge and Data Engineering, 1994, 6(1): 64-71.