

基于模糊支持向量机的单向变速决策模型研究

尤永健, 缪 芸, 倪伊楠, 李 雷

(南京邮电大学 非结构化数据计算与应用研究中心, 江苏 南京 210023)

摘 要: 汽车的智能驾驶决策问题是近年来机器学习领域的研究热点之一。驾驶员在驾驶过程中, 通常会受到多种环境因素的影响, 倘若驾驶员无法快速处理这些信息并做出正确的判断, 无疑会引起安全事故。为此, 在合理简化的前提下, 基于驾驶决策行为以及影响驾驶决策相关环境因素的研究分析, 将模糊理论和软间隔支持向量机理论相结合, 提出了基于模糊支持向量机的单向变速决策模型。该模型旨在针对实时周边环境, 为车辆驾驶员提供变速决策建议, 帮助驾驶员更加高效地制定出安全的驾驶策略。仿真实验结果表明, 相较于传统的支持向量机模型, 建立于隶属函数之上的模糊支持向量机模型具有较高的决策正确率和较好的可拓展性, 可为车辆驾驶员提供合理、安全的决策建议。

关键词: 单向变速决策模型; 模糊理论; 软间隔支持向量机; 隶属函数

中图分类号: O231

文献标识码: A

文章编号: 1673-629X(2017)07-0190-04

doi: 10.3969/j.issn.1673-629X.2017.07.042

Investigation on Single-track Variable Speed Decision Model Based on Fuzzy Support Vector Machine

YOU Yong-jian, MIAO Yun, NI Yi-nan, LI Lei

(Non-structural Data Computing and Application Research Center, Nanjing University of
Posts and Telecommunications, Nanjing 210023, China)

Abstract: Automatic driving decision-making problem is one of the hot spots in the field of machine learning in recent years. The drivers, in the driving process, are usually affected by many environmental factors. And if a driver cannot quickly deal with the information and make the right decisions, he would undoubtedly suffer huge security risks. In order to solve this problem, on the analysis of the driving decision-making behavior and related environmental factors affecting driving decision-making, a single-track variable speed decision model based on fuzzy support vector machine has been proposed in combination with fuzzy theory and soft interval support vector machine theory, which aims to provide drivers with decision-making suggestions for the current situation and to efficiently help drivers to develop a safe driving strategy. The simulation results show that compared with the traditional support vector machine, the proposed model has higher decision accuracy and relatively good expansibility suitable to provide drivers reasonable and safe recommendations.

Key words: single-track variable speed decision model; fuzzy theory; soft interval support vector machine; membership functions

0 引 言

不同驾驶行为模式的决策机制和车辆运行模式的执行机制, 是仿真多源信息刺激下驾驶员任务集聚、综合认知活动的两个重要环节^[1]。在现实生活中, 驾驶员在驾驶时通常会受到不同环境信息的影响, 而高效、正确地处理这些环境信息并做出正确的判断是避免发生交通事故的关键^[2]。

Vapnik 等提出的支持向量机(SVM)模型^[3]是在结构风险最小化原则和统计学习理论的 VC 维理论基

础上专门研究有限样本预测的机器学习方法, 能够较好地处理分类和回归问题, 具有很强的泛化能力^[4-6]。但 SVM 还存在一定的局限, 例如训练样本中存在大量的模糊信息, 导致 SVM 对孤立点或噪声的过分敏感, 容易出现“过学习”现象, 使支持向量机的性能受到影响。近年来, 许多专家学者把模糊理论^[7]和支持向量机的优点相结合, 对模糊支持向量机(Fuzzy Support Vector Machine, FSVM)进行研究^[8-9]。

基于模糊支持向量机的单向变速决策模型在一定

收稿日期: 2016-04-03

修回日期: 2016-08-14

网络出版时间: 2017-06-05

基金项目: 国家自然科学基金资助项目(61501251, 61373137, 61071167); 江苏省普通高校研究生科研创新计划资助项目(KYZZ15_0236); 南京邮电大学引进人才科研启动基金资助项目(NY214191)

作者简介: 尤永健(1995-), 男, 研究方向为机器学习及最优化理论; 李 雷, 博士, 教授, 研究方向为机器学习及最优化理论。

网络出版地址: <http://cnki.net/kcms/detail/61.1450.TP.20170605.1506.006.html>

程度上继承了模糊理论和支持向量机的优点,具有优秀的学习性能和泛化能力,能较好地解决小样本、非线性、高维数和局部极小点等实际问题,针对样本中存在的模糊信息,极大地削弱了噪声或孤立点对分类的影响,在进行驾驶行为的决策上具有较高的可信度^[10]。

为帮助驾驶员更加高效地制定出安全的驾驶策略,将智能决策行为简化为车辆的加减速行为,引入模糊理论建立起单车道变速决策模型,在真实情况下对车速、车距等数据进行采集,并进行仿真实验以实现驾驶行为的预测。

1 模糊支持向量机

在实际工程应用中,由于受到干扰、环境的影响,样本可能存在一定的模糊性,使传统 SVM 的性能受到影响。针对这种情况,Lin 提出了基于样本点模糊的 FSVM^[4],对每一个样本 x_j 设定一个隶属度 μ_j ($0 < \mu_j < 1$)。设样本训练集为 (x_1, y_1, μ_1) , (x_2, y_2, μ_2) , \dots , (x_l, y_l, μ_l) , 分类优化问题为:

$$\min \frac{1}{2} \|\omega\|^2 + C \sum_{j=1}^l \mu_j \xi_j$$

s. t.
$$\begin{cases} y_j((\omega \cdot x_j) + b) + \xi_j \geq 1 \\ \xi_j \geq 0, j = 1, 2, \dots, l \end{cases}$$

(1)

其中, C (惩罚系数)用来控制分类间隔和错分样本的个数; μ_j (隶属度)用来描述样本隶属于某类的程度; ξ_j (松弛变量)用来衡量样本被错分的度量; $\mu_j \xi_j$ 表示不同重要性的样本被错分的程度。

对于上述二次优化问题(1)的求解,运用 Lagrange 优化方法,其具体的求解过程可分为以下三步:

(1) 构造 Lagrange 函数并求其最小值:

$$L(\omega, b, \xi, \alpha, \beta) = \frac{1}{2} \|\omega\|^2 + C \sum_{j=1}^l \mu_j \xi_j - \sum_{j=1}^l \alpha_j (y_j((\omega \cdot x_j) + b) - 1 + \xi_j) - \sum_{j=1}^l \beta_j \xi_j$$

$$\alpha_j \geq 0, \beta_j \geq 0, j = 1, 2, \dots, l$$

(2)

根据 Wolfe 对偶定义,对 Lagrange 函数分别关于 ω, b, ξ_j 求极小,即:

$$\frac{\partial L(\omega, b, \xi, \alpha, \beta)}{\partial \omega} = \omega - \sum_{j=1}^l \alpha_j y_j x_j = 0$$

(3)

$$\frac{\partial L(\omega, b, \xi, \alpha, \beta)}{\partial b} = - \sum_{j=1}^l \alpha_j y_j = 0$$

(4)

$$\frac{\partial L(\omega, b, \xi, \alpha, \beta)}{\partial \xi_j} = \mu_j C - \alpha_j - \beta_j = 0$$

(5)

(2) 将式(3)~(5)转化为其对偶形式并求解。

综合考虑式(2)~(5),写出二次优化问题(1)的对偶形式:

$$\max \sum_{j=1}^l \alpha_j - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j y_i y_j K(x_i \cdot x_j)$$

s. t.
$$\begin{cases} \sum_{j=1}^l y_j \alpha_j = 0 \\ 0 \leq \alpha_j \leq \mu_j C, j = 1, 2, \dots, l \end{cases}$$

(6)

求解此对偶问题(6)可得其最优解 α^* 。

(3) 得出模糊支持向量机最优判别函数:

$$f(x) = \text{sgn}[(\omega^* \cdot x) + b^*], x \in R^n$$

(7)

其中, $\omega^* = \sum_{j=1}^l \alpha_j^* y_j x_j$, $b^* = y_i - \sum_{j=1}^l y_i \alpha_j^* (x_i x_j)$ 。

参数 C 是自定义的惩罚因子,控制着对错分样本惩罚的程度, C 越大,对错分样本的约束越大,得到分类面的间隔就越小。在该 FSVM 中,由于 μ_j 的存在, α_j 可能会对应于传统 SVM 不同类的支持向量,能有效减少噪声和孤立点对分类的影响。

2 单向变速决策模型的建立

将驾驶行为进行简化,仅考虑在单向高速公路上的驾驶场景(见图1),针对现实生活中最常见的加减速决策进行讨论分析。因此假设所处的驾驶环境为单向单车道的高速公路,所进行的决策仅与车速变化有关,在此前提下通过进一步建立模型进行智能决策。



图1 车辆驾驶实际情况

2.1 决策影响因子的选取

基于理论知识和经验知识,影响驾驶员车速控制的主要因素为与前后车之间的实际车距和与前后车之间的相对车速。实际车距和相对车速的变化必然会引起驾驶员对车速控制做出相应的决策,包括加速、减速和维持原车速。为此,建立以下五个无量纲量用以描述上述两个影响因素^[11]。五个量的详细信息如下:

- (1) X_1 :理想车速 (v_{ideal}) 与本车车速 (v_0) 之比,即 $X_1 = \frac{v_{ideal}}{v_0}$ 。
- 当 $X_1 < 1$ 时,车速相对较快,可适当减速;
- 当 $X_1 = 1$ 时,速度相对适中,可维持原车速;

当 $X_1 > 1$ 时,速度相对较慢,可适当加速。

(2) X_2 : 前车车速 (v_b) 与本车车速 (v_0) 之比,即

$$X_2 = \frac{v_b}{v_0}。$$

当 $X_2 < 1$ 时,相对于前车车速较快,可适当减速;

当 $X_2 = 1$ 时,与前车车速相当,可维持原车速;

当 $X_2 > 1$ 时,相对于前车车速较慢,可适当加速。

(3) X_3 : 后车车速 (v_a) 与当前车速 (v_0) 之比,即

$$X_3 = \frac{v_a}{v_0}。$$

当 $X_3 < 1$ 时,相对于后车车速较快,可适当减速;

当 $X_3 = 1$ 时,与后车车速相当,可维持原车速;

当 $X_3 > 1$ 时,相对于后车车速较慢,可适当加速。

(4) X_4 : 和前车的实际距离 (d_b) 与安全车距

(d_0) 之比,即 $X_4 = \frac{d_b}{d_0}。$

当 $X_4 < 1$ 时,与前车距离较小,相对不安全,可适当减速;

当 $X_4 \geq 1$ 时,与前车距离较为合适,相对安全,适当加速或维持原车速均可。

(5) X_5 : 和后车的实际距离 (d_a) 与安全车距

(d_0) 之比,即 $X_5 = \frac{d_a}{d_0}。$

当 $X_5 < 1$ 时,与后车距离较小,相对不安全,可适当减速;

当 $X_5 \geq 1$ 时,与后车距离较为合适,相对安全,适当加速或维持原车速均可。

另外,将驾驶员针对以上情况所可能采取的决策定义为: S_1 —车辆加速; S_2 —车辆减速; S_3 —车速不变。

2.2 隶属度的计算

在确定了影响驾驶决策的五个量后,需要对它们之间的内在联系进行分析。

根据专家经验,可以给出以下 9 条模糊推理规则^[11]:

规则 1: 如果 X_1 倾向于加速,且 X_4 允许加速,则采取决策 S_1 。

规则 2: 如果 X_1 倾向于加速,但 X_4 不具备加速条件,则采取决策 S_3 。

.....

规则 9: 如果 X_1 倾向于匀速,且无来自前向 X_2 和后向 X_3 的加/减速压力,则采取决策 S_3 。

根据以上 9 条推理规则可以看出,一般情况下,车辆驾驶员所采取的车速变化决策并非仅仅取决于单一影响因素,而是需要分析各因素的重要程度并综合考虑五个因素以做出合理的决策。

因此,构造权重矩阵 $\mathbf{W} = (w_{ij})_{5 \times 5}$,用以描述在不

同情况下各因子对某一决策隶属度的权重。其中,行向量 \mathbf{w}_i 表示采取决策模式为 i 的情况下,该因子对决策隶属度的权重。

由此,构造计算隶属度的公式为^[7]:

$$\mu = M\left(\sum_{i=1}^5 \mathbf{w}_i * \mathbf{x}_i - c\right) \tag{8}$$

其中, c 表示样本中心,在该模型中,定义 c 为 1。

对于隶属函数 $M(x)$ ^[7], 选用 S 型分段函数,其数学表达式如下:

$$M(x) = \begin{cases} 1 & x > a \\ 1 - 2\left(\frac{x - c}{c - a}\right)^2 & a \leq x \leq b \\ 2\left(\frac{x - c}{c - a}\right)^2 & b \leq x \leq c \\ 0 & x > c \end{cases} \tag{9}$$

其中, a, b, c 均为参数,且满足 $b = \frac{a + c}{2}$ 。特别

地,当 $x = b$ 时, $M(x) = 0.5$

2.3 核函数的选取

在得出样本所对应的隶属度后,选取 RBF 核函数和线性核函数^[12-14]作为模糊支持向量机的核函数进行对比实验。

RBF 核函数: $k(x, y) = \exp(-\gamma \|\mathbf{x} - \mathbf{y}\|^2)$,
 $\gamma > 0$ 。

线性核函数: $k(x, y) = \mathbf{x}^T \mathbf{y} + c$ 。

3 实 验

3.1 实验过程

(1) 数据采集: 在测试车辆前后端均安装测速雷达和激光测距仪,用以采集前后车辆的动态信息,包括驾驶车辆本身瞬时车速 (v_0)、与之相邻的前后车辆的瞬时车速 (v_b 、 v_a)、与前后车辆的瞬时距离 (d_b 、 d_a) 等五个量;另外,配备了多个实时摄像头,用来记录车辆驾驶员针对不同情况所采取的实际决策。

共采集了 6 000 组数据,并对其做无量纲化处理,如表 1 所示,列出前二十组。

表 1 实际车辆动态信息与驾驶员的实际决策行为

理想速 /瞬时	前车速 /瞬时	后车速 /瞬时	前车距 /安全	安全 /后车距	是否 变速	如何 变速
1.148 4	0.909 8	1.447 1	0.615 9	0.379 6	1	-1
1.317 6	1.219 8	1.186 8	0.705 6	0.494 5	1	-1
0.728 1	1.033 7	1.246 7	0.542 3	0.982	1	-1
1.105 7	1.356	0.792 2	0.704 4	1.189 9	-1	0
0.477 1	0.755 9	1.041 7	1.116 3	1.248 4	1	-1
1.220 4	1.13	0.645 1	0.924 1	1.138 3	-1	0
0.645 1	1.122 4	1.075 8	1.184 4	0.943 1	1	1
1.255	0.480 1	1.361 2	1.325	1.167 6	1	1

续表 1

理想速 /瞬时	前车速 /瞬时	后车速 /瞬时	前车距 /安全	安全 /后车距	是否 变速	如何 变速
1.137 1	1.093 7	0.651 9	1.114 7	1.788 3	-1	0
0.869 9	1.550 1	0.933 9	1.245 2	1.625 1	1	1
0.532	0.621 2	0.536 4	1.238 7	0.981 2	1	-1
1.421 1	0.798 5	1.191 7	1.745 4	1.447 3	1	1
1.048 3	0.678 5	1.389 2	1.239 4	0.803 8	-1	0
1.530 4	1.357	1.361 6	0.528 2	1.110 9	1	1
1.002 7	0.824	0.675 7	1.093 5	1.045 7	-1	0
0.437 6	1.237 6	0.775 8	1.211 1	1.241 4	1	1
1.534 4	0.479 9	0.967 7	1.298 1	0.615 8	1	-1
0.907 9	0.423 7	0.875 1	1.258 4	1.197 4	1	-1
1.292 3	1.315 4	0.833 2	0.844 4	1.053 3	-1	0
0.771 6	1.187 6	0.201 7	1.008 8	1.104	1	-1

(2)模型训练:对于采集的6 000组数据,每次选出300组进行实验。每次实验选取其中283组作为训练样本。在训练过程中,首先计算出它们关于相应决策的隶属度,然后用其对模糊支持向量机模型进行训练。

(3)模型预测:运用训练好的模糊支持向量机模型对单次训练剩余17种车辆动态状况下应当采取的驾驶决策行为进行预测。将预测结果与驾驶员的实际决策进行对比,分析模型的优劣。

3.2 实验结果及分析

按照上述实验过程,一共进行了20次实验,分别计算所提出的模型在20次实验中的决策正确率。同时,基于相同的数据又对传统的支持向量机进行训练。

分别选用RBF和线性核函数,对传统支持向量机的预测正确率与所提出模型的正确率做两次对比,如图2、图3所示。

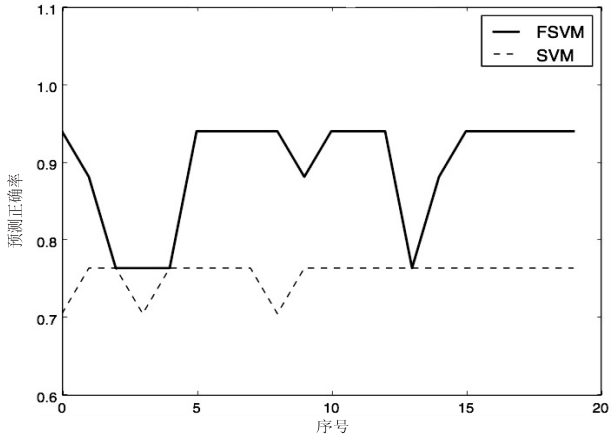


图2 RBF核函数下支持向量机和模糊支持向量机的预测对比

通过图2和图3可以看出,采用RBF核函数相较于线性核函数的模型来说具有较高的稳定性,其预测结果多集中在数据数值附近,波动较小;而采用线性核

函数预测的成功率均出现了较大范围的波动。从训练速度上看,使用RBF核函数的模型训练速度快于使用线性核函数的模型数倍。可见在实际应用中,RBF核函数更适合于所提出的模型。从模糊支持向量机与支持向量机的预测成功率可以看出,无论是采用了RBF核函数还是采用了线性核函数,模糊支持向量机模型的预测结果与驾驶员的决策相似度均高于传统支持向量机,并且RBF核函数下的模糊支持向量机的预测成功率提升十分明显。

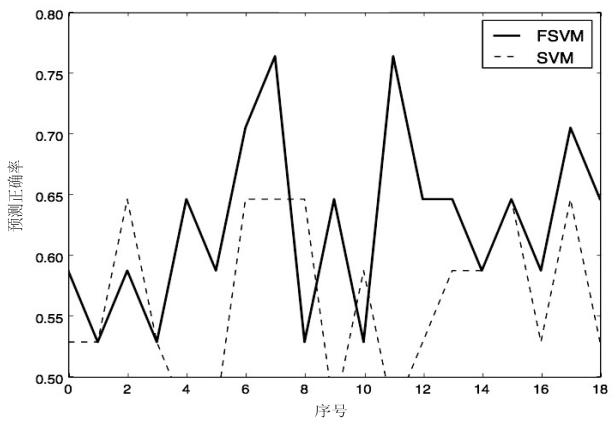


图3 线性核函数下支持向量机和模糊支持向量机的预测对比

因此,所建立的单车道变速决策模型在一定的训练后预测得出的结果是高度可用的,其在智能驾驶行为的预测上是成功的。

4 结束语

为帮助驾驶员更加高效地制定出安全的驾驶策略,在智能驾驶领域已有算法的基础上进行了算法改进,在将智能决策行为简化为车辆的加减速行为的前提下,引入模糊理论建立了单车道变速决策模型,该模型基于实际采集的车辆信息对驾驶行为进行预测。实验结果表明,该模型有效可行,不仅可以较为准确地反映出驾驶过程中环境信息与驾驶员所采取驾驶决策的映射关系,而且相较于传统支持向量机模型有更高的优越性。

需要说明的是,实际驾驶行为不仅仅只有加减速两种,还存在是否变道、是否转弯、是否掉头等更为复杂的决策,因此模型还需要针对实际驾驶情况进行进一步的研究改进。

参考文献:

[1] 史忠植. 知识发现[M]. 北京:清华大学出版社,2011.
[2] 王 强,沈永平,陈英武. 基于支持向量机的智能决策方法[J]. 系统工程,2005,23(10):111-116.
[3] Vapnik V. The nature of statistical learning theory[M]. [s.
(下转第199页)

上述实验可以说明,通过提出的 PTF-IDF 算法得到的领域情感词进而构建领域情感词典,并代入各领域购物评论进行情感分类,能够有效地提升分类准确率,从而验证了该算法的有效性。

6 结束语

面对如今网上购物商城中,用户对消费购物体验过程给出的海量评论,如何根据评论内容较为迅速地提炼出有价值的信息,并对评论的情感倾向做出正确分类,已成为目前迫切需要解决的问题。针对不同领域之间词语相似度存在不同,在文本中表达意义存在差异的特点,通过扩展基础情感词典,利用采集的语料中不同词性词语对评论文本情感倾向性的影响程度筛选合适的词语作为候选情感词,结合评论候选情感词与不同领域的文本之间关联程度信息,构建购物评论领域情感词典,代入购物评论的情感分类研究中。实验结果表明,该方法能有效提高分类准确度,并降低情感特征空间维数,灵活适应情感词在各个领域间存在的语义差异问题。

下一步的研究工作主要是利用现有语料资源和中文语法规则动态地修正词典,达到规范完善领域词典的目的;同时,综合考虑主题词语、评论对象对情感倾向计算的影响,结合上下文相关信息,提供一种更加合理全面的情感倾向计算方式。

参考文献:

- [1] 薛益定. 中文情感分析研究综述[J]. 电脑编程技巧与维护,2016(5):22-24.
- [2] Neviarouskaya A, Prendinger H, Ishizuka M. SentiFul: a lexicon for sentiment analysis[J]. IEEE Transactions on Affective Computing, 2011, 2(1):22-36.
- [3] 周俊臣, 赵 晖. WEB 文本情感倾向性分析研究综述[J]. 计算机光盘软件与应用, 2014, 17(1):303-304.

(上接第 193 页)

- [1] :Springer Science & Business Media,2013.
- [4] Joachims T. Making large-scale SVM learning practical[R]. [s.l.]:[s.n.],1998.
- [5] Hsu C W, Chang C C, Lin C J. A practical guide to support vector classification[EB/OL]. 2010-04-15. <https://www.cs.sfu.ca/people/Faculty/teaching/726/spring11/svmguide.pdf>.
- [6] 张学工. 关于统计学习理论与支持向量机[J]. 自动化学报,2000,26(1):32-42.
- [7] Zadeh L A. Fuzzy sets[J]. Information & Control, 1965, 8(3):338-353.
- [8] Lin C F, Wang S D. Fuzzy support vector machines[J]. IEEE Transactions on Neural Networks, 2002, 13(2):464-471.

- [4] 饶洋辉,李 青,刘文印,等. 公众文本之情感词典研究进展[J]. 中国科学:信息科学,2014,44(7):825-835.
- [5] 张成功,刘培玉,朱振方,等. 一种基于极性词典的情感分析方法[J]. 山东大学学报:理学版,2012,47(3):47-50.
- [6] 王宝成,何新宇. 基于改进情感词域识别的舆情情感分析研究[J]. 电子技术与软件工程,2016(3):167.
- [7] 刘清松,张仰森. 基于多词典融合的词汇语义倾向判别[J]. 计算机技术与发展,2015,25(5):104-109.
- [8] 李爱萍,邸 鹏,段利国. 基于句子情感加权算法的篇章情感分析[J]. 小型微型计算机系统,2015,36(10):2252-2256.
- [9] Pang B, Lee L, Vaithyanathan S. Thumbs up? sentiment classification using machine learning techniques[C]//Proceedings of the conference on empirical methods in natural language processing. Philadelphia:ACL Press,2002:79-86.
- [10] Yu H, Hatzivassiloglou V. Towards answering opinion questions: separating facts from opinions and identifying the polarity of opinion sentences[C]//Proceedings of the EMNLP 2003. Morristown:ACL Press,2003:129-136.
- [11] 樊康新. 基于 SVM 的网络文本情感分类系统的设计[J]. 计算机时代,2015(12):34-37.
- [12] 蒋婉婷,孙 蕾,钱 江. 基于 Hadoop 的朴素贝叶斯算法在中文微博情感分类中的研究与应用[J]. 计算机应用与软件,2015,32(7):60-62.
- [13] 张 森,曹 晖. 基于《知网》概念定义的情感词典构建研究[J]. 计算机工程与应用,2015,51(17):118-123.
- [14] Lawrence P, Sergey B. The anatomy of a large scale hyper textual web search engine[J]. Computer Networks & ISDN Systems, 2012, 56(18):3825-3833.
- [15] 顾益军,刘小明. 融合多种情感资源的微博情感分类研究[J]. 计算机科学,2015,42(4):209-212.
- [16] 陈 明. 多极性形容词的倾向判别及其应用研究[D]. 郑州:郑州大学,2012.
- [17] 李宪毅,刘培玉,朱振方,等. 基于改进情感关键词抽取的自动情感摘要技术[J]. 计算机工程与设计,2016,37(3):778-782.

- [9] Inoue T, Abe S. Fuzzy support vector machines for pattern classification[C]//International joint conference on neural networks. [s.l.]:IEEE,2001:1449-1454.
- [10] 陈雪梅,田 赓,苗一松. 面向智能驾驶行为的机器学习[J]. 道路交通与安全,2014,14(6):60-64.
- [11] 王伟钰,曲仕茹. 基于模糊神经网络的自动驾驶决策系统研究[J]. 计算机测量与控制,2009,17(9):1711-1713.
- [12] Smola A J. Learning with kernels I support vector machines[J]. Lecture Notes in Computer Science, 2008, 42(4):1-28.
- [13] 奉国和. SVM 分类核函数及参数选择比较[J]. 计算机工程与应用,2011,47(3):123-124.
- [14] 李盼池,许少华. 支持向量机在模式识别中的核函数特性分析[J]. 计算机工程与设计,2005,26(2):302-304.