

Fig.2 Comparison of the response time of the four algorithms in census data set
图2 Census 数据集上 4 种算法的响应时间比较

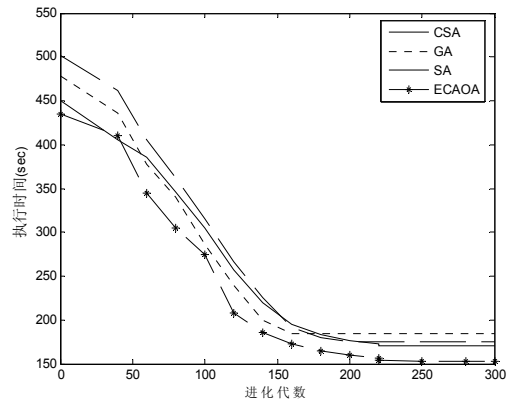


Fig.3 Comparison of the execution time of the four algorithms in census data set
图3 Census 数据集上 4 种算法的执行时间比较

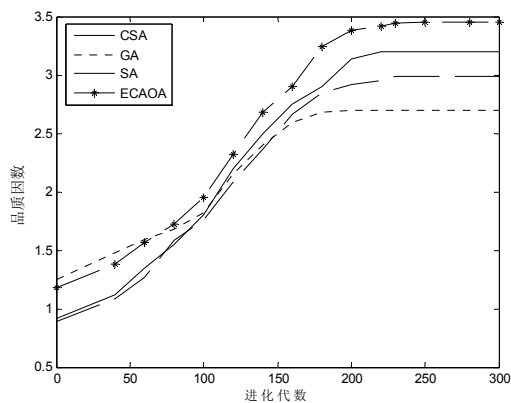


Fig.4 Comparison of the quality factor of the four algorithms in census data set
图4 Census 数据集上 4 种算法的品质因数比较

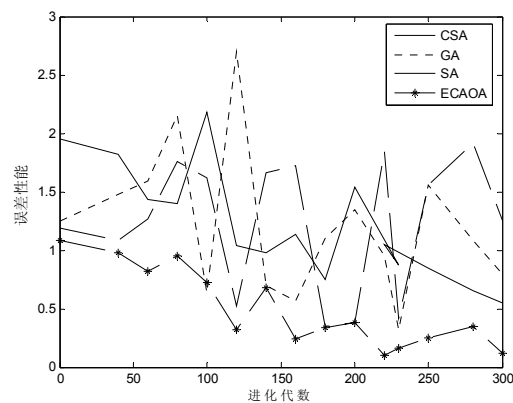


Fig.5 Comparison of the error performance of the four algorithms in census data set
图5 Census 数据集上 4 种算法的误差性能比较

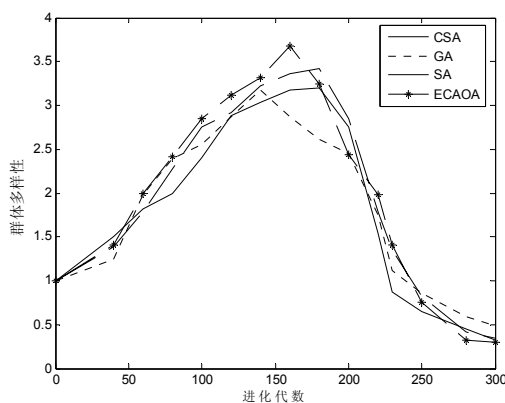


Fig.6 Comparison of the population diversity of the four algorithms in census data set
图6 Census 数据集上 4 种算法的群体多样性比较

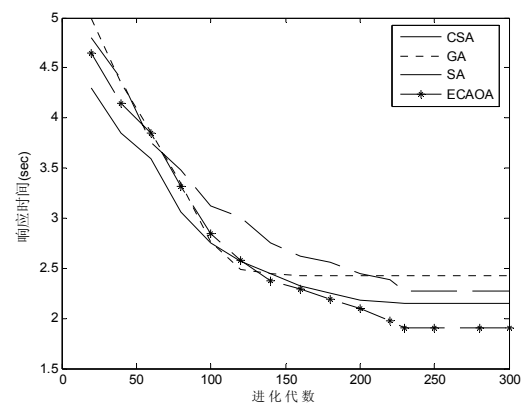


Fig.7 Comparison of the response time of the four algorithms in chess data set
图7 Chess 数据集上 4 种算法的响应时间比较

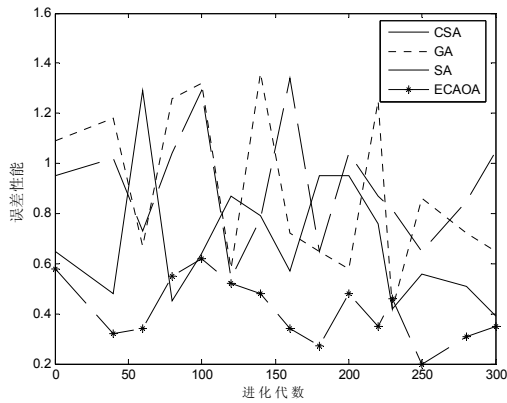


Fig.8 Comparison of the execution time of the four algorithms in chess data set

图8 Chess数据集上4种算法的执行时间比较

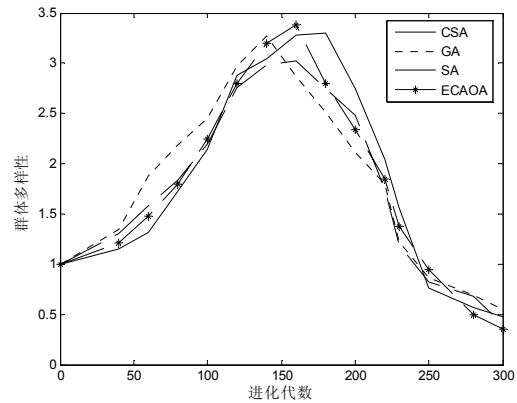


Fig.9 Comparison of the quality factor of the four algorithms in chess data set

图9 Chess数据集上4种算法的品质因数比较

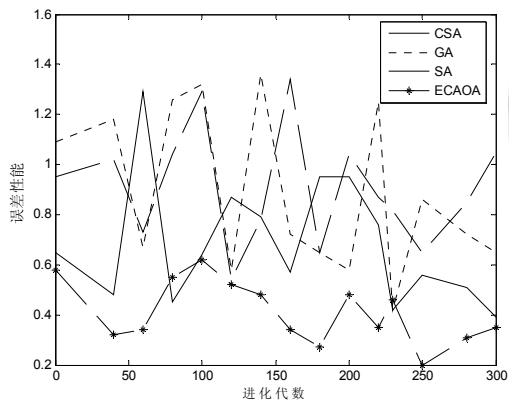


Fig.10 Comparison of the error performance of the four algorithms in chess data set

图10 Chess数据集上4种算法的误差性能比较

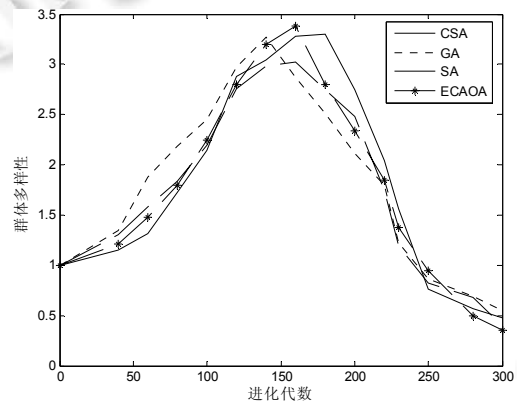


Fig.11 Comparison of the population diversity of the four algorithms in chess data set

图11 Chess数据集上4种算法的群体多样性比较

通过第1类实验,我们不难分析出,ECAOA算法在响应时间、执行时间、品质因数、误差性能和群体多样性这5方面都表现出良好的性能.这主要是因为ECAOA算法在变异和交叉等操作产生新的抗体后,首先通过模拟退火的Metropolis准则来决定新抗体是否保留,从而有效保证新的抗体具有更好的性能.ECAOA算法将CSA算法和SA算法的优点进行结合,有效避免了陷入局部极值,能够有效改良算法的品质因数、误差性能和群体多样性等性能指标参数.在进化前期,ECAOA算法的性能与其他3种算法相比并没有表现出明显的优势,相反,在局部情况下,其性能还比较落后.导致这种情况的原因是,ECAOA算法在进行首次选择操作时要计算每个抗体的增强度,从而增大了所需的计算量.CSA算法在绝大多数情况下的性能要优于SA算法和GA算法,主要是克隆操作不仅扩大了解的空间范围,而且提高了优良抗体的比例,这样将积极影响到群体多样性和误差性能等性能指标参数.GA算法在大多数情况下的性能表现不如其他3种算法,主要是因为该算法很容易出现早熟收敛现象,导致其群体多样性较差,误差性能较大.

为了更好地评估ECAOA算法的性能,本节结合算法的执行时间、挖掘的关联规则数目和关联规则准确率对算法的挖掘质量进行如下定义:

定义 24. 如果算法挖掘的关联规则数目用 $total$ 表示,算法的执行时间用 t_e 表示,关联规则的准确率用 θ 表

示,则算法的挖掘质量 ϕ 描述如下:

$$\phi = \frac{total \times \theta}{t_e} \tag{25}$$

从公式(25)中不难分析出:执行时间 t_e 越少,关联规则数目 $total$ 越大,准确率 θ 越高,则算法的挖掘质量 ϕ 越大,算法的整体性能越优良.

第 2 类实验主要从运行时间、关联规则数目、准确率和挖掘质量这 4 个方面进行比较,在 census 数据集上的实验结果如图 12~图 15 所示.

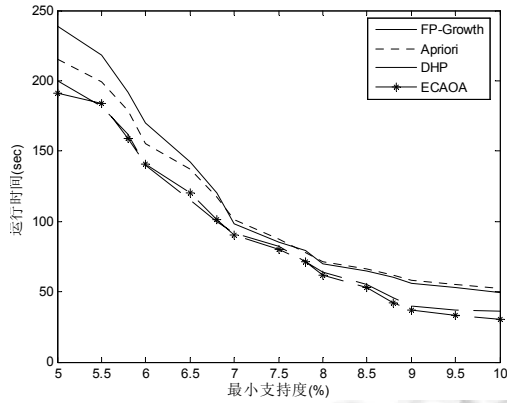


Fig.12 Comparison of the execution time of the four algorithms in census data set

图 12 Census 数据集上 4 种算法的运行时间比较

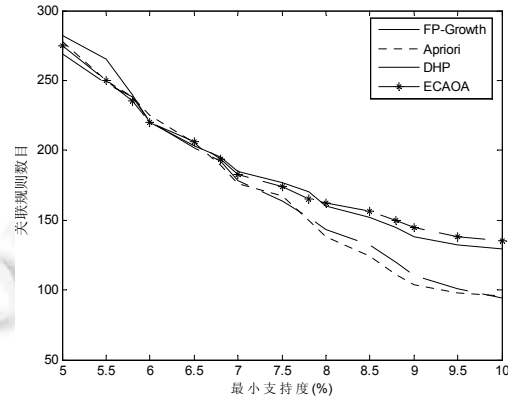


Fig.13 Comparison of number of association rules of the four algorithms in census data set

图 13 Census 数据集上 4 种算法的关联规则数目比较

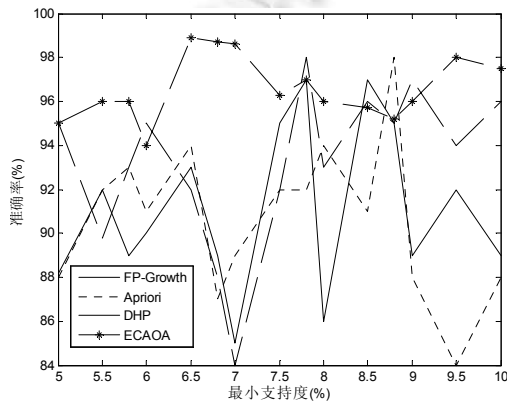


Fig.14 Comparison of the accuracy rate of the four algorithms in census data set

图 14 Census 数据集上 4 种算法的准确率比较

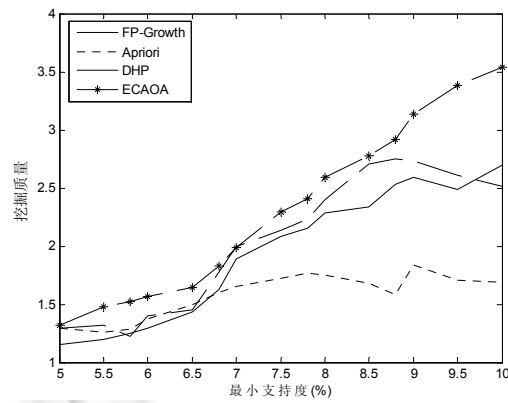


Fig.15 Comparison of the mining quality of the four algorithms in census data set

图 15 Census 数据集上 4 种算法的挖掘质量比较

在图 12 的执行时间对比实验中,当最小支持度小于 7.8%时,ECAOA 算法和 DHP 算法的性能非常接近,Apriori 算法和 FP-Growth 算法的性能表现得相对较差;当最小支持度大于 7.8%时, ECAOA 算法的优势最明显,而 Apriori 算法和 FP-Growth 算法的执行时间几乎相等,其运行时间比 DHP 算法大约要少 10s,比 Apriori 算法和 FP-Growth 算法大约要少 28s.在图 13 中,当最小支持度在 5.75%~7%范围时,4 种挖掘算法得到的关联规则数目非常接近;当最小支持度大于 7%时,ECAOA 算法和 DHP 算法都表现出良好的性能,其挖掘出的关联规则数目比 DHP 算法大约多 5.07%,比 FP-Growth 算法和 Apriori 算法大约多 25.23%.

在图 14 中,ECAOA 算法的准确率平均达到 95.88%,FP-Growth,Apriori 和 DHP 算法的准确率平均分别为 90.39%,94.08%,93.92%。在图 15 显示的挖掘质量对比中,ECAOA 算法的优势表现得非常明显,随着最小支持度的逐渐变大,ECAOA 算法的挖掘质量明显提高。

在 chess 数据集上的实验结果如图 16~图 19 所示。

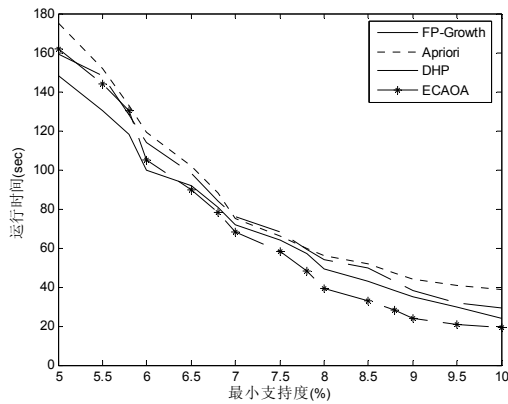


Fig.16 Comparison of the execution time of the four algorithms in chess data set

图 16 Chess 数据集上 4 种算法的运行时间比较

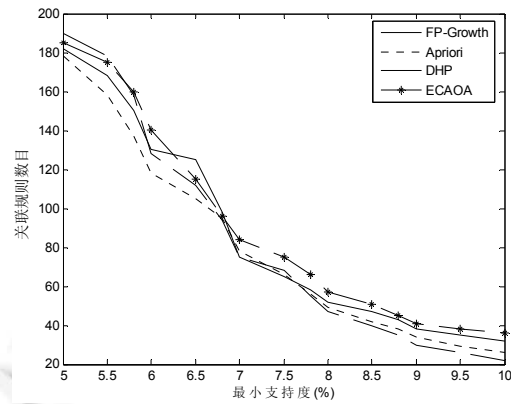


Fig.17 Comparison of number of association rules of the four algorithms in chess data set

图 17 Chess 数据集上 4 种算法的关联规则数目比较

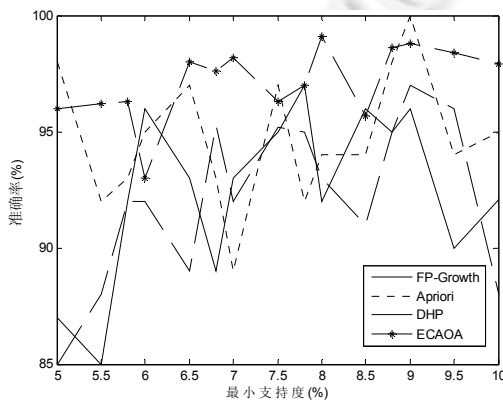


Fig.18 Comparison of the accuracy rate of the four algorithms in chess data set

图 18 Chess 数据集上 4 种算法的准确率比较

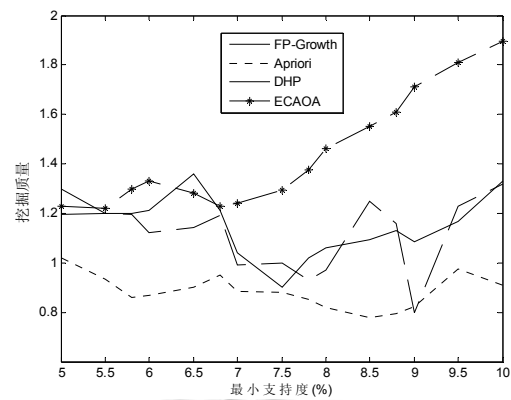


Fig.19 Comparison of the mining quality of the four algorithms in chess data set

图 19 Chess 数据集上 4 种算法的挖掘质量比较

在图 16 的运行时间对比实验中,当最小支持度小于 6.5%时,ECAOA 算法与 Apriori 算法和 DHP 算法三者表现出来的性能互有优劣,但是 FP-Growth 算法的性能表现得相对较好;当最小支持度大于 6.5%时,ECAOA 算法的优势最明显,随着最小支持度逐渐变大,ECAOA 算法的运行缓慢减少。在图 17 中,当最小支持度小于 6.8%时,ECAOA 算法的性能稍微优于 Apriori,DHP,FP-Growth 算法;当最小支持度大于 6.8%时,ECAOA 算法表现出优良的性能,其挖掘出的关联规则数目比 FP-Growth 算法大约多 9.38%,比 Apriori 算法大约多 21.45%,比 DHP 算法大约多 23.68%。

在图 18 中,ECAOA 算法的准确率平均达到 97.05%,FP-Growth,Apriori 和 DHP 算法的准确率平均分别为 93.12%,94.67%,90.32%。在图 19 显示的挖掘质量对比中,随着最小支持度的不断变大,ECAOA 算法的挖掘质量逐渐提高。当最小支持度大于 6.8%时,ECAOA 算法的性能最优良,Apriori 算法的性能表现最差,DHP 算法的性能

最不稳定.

通过第 2 类实验,我们不难分析出:ECAOA 算法在绝大多数情况下,在运行时间、关联规则数目、准确率这 3 方面的性能都要稍微超过其他 3 种挖掘算法,在挖掘质量方面则明显优于其他 3 种挖掘算法.这主要是因为 ECAOA 算法充分结合了克隆选择算法和模拟退火算法的优点,在每次挖掘过程中都采取并行的方式进行搜索,同时在算法的执行过程,总是将优良的抗体在群体中的规模不断扩大,变异和交叉操作又不断地丰富了种群的多样性,并且将生存度最高的抗体通过克隆抑制操作直接保留到下一代,同时,种群更新操作将性能较差的抗体直接进行删除.ECAOA 算法在整个数据挖掘期间充分体现出人工免疫算法所具备的并行性和鲁棒性特征.Apriori 算法在大多数情况下的性能表现都不如 DHP 和 FP-Growth 算法,主要是因为其挖掘的规则存在大量冗余,并且算法需要多次扫描数据库,导致所消耗的时间较大,而 DHP 算法通过使用 TID,避免了反复读取事务数据库,FP-Growth 算法将频繁项集的数据库压缩成一棵频繁模式树的方式,能够有效提高算法的执行效率.

通过应用 ECAOA,CSA,GA,SA,Apriori,DHP 和 FP-Growth 算法对 census 和 chess 数据集从不同方面进行测试,实验结果显示了本文设计的 ECAOA 算法具有较快的执行时间和较强的搜索能力,能够有效地避免 CSA 算法可能存在的早熟收敛现象,同时具有良好的品质因数、挖掘质量和群体多样性,能够成功地应用于数据挖掘领域中.

7 结束语

针对传统的克隆选择算法可能存在的早熟收敛现象和缺少交叉操作问题,提出一种高效的克隆优化退火算法 ECAOA.它结合了模拟退火算法与免疫系统的克隆选择机制,并根据退火选择概率进行选择操作,同时,根据模拟退火的 Metropolis 接受准则设计了交叉操作,不仅能够有效保持全局搜索和局部搜索的平衡,而且可以有效提高算法的搜索效率,从而加快算法的收敛速度.在第 1 类数据挖掘实验测试中,ECAOA 算法在响应时间、执行时间、品质因数、误差性能和群体多样性这 5 个方面与 CSA,SA,GA 算法相比都表现出较好的性能.在第 2 类数据挖掘实验测试中,ECAOA 算法在绝大多数情况下,在运行时间、关联规则数目、准确率这 3 个方面的性能都要稍微超过其他 3 种挖掘算法,在挖掘质量方面,则明显优于其他 3 种挖掘算法,并解决了传统的克隆选择算法可能存在的早熟收敛问题.

References:

- [1] Mohammadi M, Raahemi B, Akbari A, Nassersharif B, Moeinzadeh H. Improving linear discriminant analysis with artificial immune system-based evolutionary algorithms. *Information Sciences*, 2012,189(4):219–232. [doi: 10.1016/j.ins.2011.11.032]
- [2] Naderi B, Khalili M, Tavakkoli-Moghaddam R. A hybrid artificial immune algorithm for a realistic variant of job shops to minimize the total completion time. *Computers & Industrial Engineering*, 2009,56(4):1494–1501. [doi: 10.1016/j.cie.2008.09.031]
- [3] Molla-Alizadeh-Zavardehi S, Hajiaghaei-Keshteli M, Tavakkoli-Moghaddam R. Solving a capacitated fixed-charge transportation problem by artificial immune and genetic algorithms with a Prüfer number representation. *Expert Systems with Applications*, 2011, 38(8):10462–10474. [doi: 10.1016/j.eswa.2011.02.093]
- [4] Aragón VS, Esquivel SC, Coello Coello CA. An immune algorithm with power redistribution for solving economic dispatch problems. *Information Sciences*, 2015,295(2):609–632. [doi: 10.1016/j.ins.2014.10.026]
- [5] Ulutas BH, Kulturel-Konak S. An artificial immune system based algorithm to solve unequal area facility layout problem. *Expert Systems with Applications*, 2012,39(5):5384–5395. [doi: 10.1016/j.eswa.2011.11.046]
- [6] Liu RC, Zhu BB, Bian RY, Ma YJ, Jiao LC. Dynamic local search based immune automatic clustering algorithm and its applications. *Applied Soft Computing*, 2015,27(2):250–268. [doi: 10.1016/j.asoc.2014.11.026]
- [7] Aydin I, Karakose M, Akin E. A multi-objective artificial immune algorithm for parameter optimization in support vector machine. *Applied Soft Computing*, 2011,11(1):120–129. [doi: 10.1016/j.asoc.2009.11.003]
- [8] Sakthivel VP, Bhuvaneshwari R, Subramanian S. Artificial immune system for parameter estimation of induction motor. *Expert Systems with Applications*, 2010,37(8):6109–6115. [doi: 10.1016/j.eswa.2010.02.034]

- [9] Turkoglu I, Didem Kaymaz E. A hybrid method based on artificial immune system and k -NN algorithm for better prediction of protein cellular localization sites. *Applied Soft Computing*, 2009,9(2):497–502. [doi: 10.1016/j.asoc.2008.07.003]
- [10] Hussain AJ, Fergus P, Al-Askar H, Al-Jumeily D. Dynamic neural network architecture inspired by the immune algorithm to predict preterm deliveries in pregnant women. *Neurocomputing*, 2015,151(3):963–974. [doi: 10.1016/j.neucom.2014.03.087]
- [11] Yu HM, Yao PJ. Combined genetic algorithm/simulated annealing algorithm for large-scale system energy integration. *Computers and Chemical Engineering*, 2000,8(24):2023–2035. [doi: 10.1016/S0098-1354(00)00601-3]
- [12] Ling J, Cao Y, Ying JH, Xu GX, Huang TX. An algorithm for Generating diverse antibody based on niche technology. *Acta Electronica Sinica*, 2003,31(8):1130–1133 (in Chinese with English abstract). [doi: 10.3321/j.issn:0372-2112.2003.08.003]
- [13] Xi GY, Yue JP, Zhou BX, Tang P. Application of an artificial immune algorithm on a statistical model of dam displacement. *Computers & Mathematics with Applications*, 2011,62(10):3980–3986. [doi: 10.1016/j.camwa.2011.09.057]
- [14] Mohammad AH, Zitar RA. Application of genetic optimized artificial immune system and neural networks in spam detection. *Applied Soft Computing*, 2011,11(4):3827–3845. [doi: 10.1016/j.asoc.2011.02.021]
- [15] Li SZ, Li MY, Pan YX. Genetic annealing algorithm and its convergence analysis. *Control Theory and Applications*, 2001,19(3):376–381.
- [16] Duan HB, Zhang XY, Xu CF. *Bio-Inspired Computing*. Beijing: Science Press, 2010. 98–105 (in Chinese).
- [17] Huang YR. *Intelligent Optimization Algorithm and Its Application*. Beijing: National Defense Industry Press, 2008. 198–210 (in Chinese).
- [18] Farzanyar Z, Kangavari M, Cercone N. Max-FISM: Mining (recently) maximal frequent itemsets over data streams using the sliding window model. *Computers & Mathematics with Applications*, 2012,64(6):1706–1718. [doi: 10.1016/j.camwa.2012.01.045]
- [19] Yang BR, Li JH, Song W, Li X. KD (D&K): A new knowledge discovery process model for complex systems. *Acta Automatica Sinica*, 2007,33(2):151–155 (in Chinese with English abstract). [doi: 10.1360/aas-007-0151]

附中文参考文献:

- [12] 凌军,曹阳,尹建华,徐国雄,黄天锡.基于小生境技术的多样性抗体生成算法. *电子学报*,2003,31(8):1130–1133. [doi: 10.3321/j.issn:0372-2112.2003.08.003]
- [16] 段海滨,张祥银,徐春芳. *仿生智能计算*.北京:科学出版社,2010. 98–105.
- [17] 黄友锐. *智能优化算法及其应用*.北京:国防工业出版社,2008. 198–210.
- [19] 杨炳儒,李晋宏,宋威,李欣.面向复杂系统的知识发现过程模型 KD(D&K)及其应用. *自动化学报*,2007,33(2):151–155. [doi: 10.1360/aas-007-0151]



舒万能(1981—),男,湖北武汉人,博士,讲师,主要研究领域为智能计算,云计算.



丁立新(1967—),男,博士,教授,博士生导师,CCF 专业会员,主要研究领域为智能计算,智能信息处理.