

差分进化算法中参数自适应选择策略研究

汪慎文^{1,2,3} 张文生² 丁立新³ 谢承旺⁴ 郭肇禄⁵

(石家庄经济学院信息工程学院 石家庄 050031)¹

(中国科学院自动化研究所 北京 100190)² (武汉大学计算机学院软件工程国家重点实验室 武汉 430072)³

(华东交通大学软件学院 南昌 330013)⁴ (江西理工大学理学院 赣州 341000)⁵

摘 要 参数选择本身是一个组合优化问题,尽管过去提出了很多方法,但是参数选择依然令人困惑,为此提出适用于差分进化算法的参数自适应选择策略。该策略在进化的过程中动态评估参数的性能,并根据其结果指导下一次迭代过程的参数选择。从参数库的建立、参数评分机制和参数配置机制 3 方面展开研究,对比实验结果表明,该方法效果良好。

关键词 差分进化算法,参数自适应,参数选择

中图分类号 TP301 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2015.11.052

Research on Parameter Self-selection Strategy of Differential Evolution

WANG Shen-wen^{1,2,3} ZHANG Wen-sheng² DING Li-xin³ XIE Cheng-wang⁴ GUO Zhao-lu⁵

(School of Information Engineering, Shijiazhuang University of Economics, Shijiazhuang 050031, China)¹

(Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China)²

(State Key Laboratory of Software Engineering, School of Computer, Wuhan University, Wuhan 430072, China)³

(School of Software, East China Jiaotong University, Nanchang 330013, China)⁴

(School of Science, Jiangxi University of Science and Technology, Ganzhou 341000, China)⁵

Abstract The selection of the parameter itself is a combinatorial optimization problem. Although a considerable number of works have been conducted, it is known to be a puzzled task. In this paper, a DE algorithm was proposed that uses a new mechanism to parameter self-selection, which dynamically learns from their previous experiences and selects the best performing combinations of parameters for the next generation during the convergence process. We firstly designed the mechanism including three aspects: building of parameter database, score of parameter performance and selection of parameter combination, then we conducted the experiments on some benchmark functions to judge the performance. The results show that the DE with the new mechanism obtains promising performance.

Keywords Differential evolution, Parameter self-adaptation, Parameter selection

1 引言

差分进化算法(Differential Evolution, DE)^[1]是由美国学者 Store 和 Price 于 1995 年提出的一种简单、高效的全局优化方法,国内也有研究者翻译为“差分演化算法”、“差异演化算法”。正是由于差分进化算法具有向种群个体学习的能力,使得其拥有其他进化算法无法相比的性能,并在诸多领域取得了惊人的效果。尽管差分进化算法性能卓越,但改善 DE 性能依然是研究的重要方向,其目的是适应更为复杂的优化问题和获得更高的求解质量。近年来,研究人员从多个角度不断改进算法以获得更有鲁棒性的差分进化算法,改进算法

主要集中在 3 方面^[2]: (1)操作算子的改进,提出一些新的参数控制方法(如 FADE^[3]、JADE^[4]、jDE^[5]等)和新设计的差分变异策略(如三角变异策略^[6]、精英变异策略^[7]等); (2)基于静态知识指导的差分进化集成算法,该类算法主要是指 DE 与其他具有某种优良特性的策略集成,如与分布评估算法集成的 DE/EDA^[8]、与反向学习策略集成的 ODE^[9]、与粒子群优化算法集成的 PSO-DE^[10]、3 个差分变异策略的同质集成的 CoDE^[11]等; (3)基于动态知识指导的差分进化集成算法,该类算法对进化过程中的不同变异策略产生的新解进行评估,从而挖掘出有用的动态知识,并指导下一个迭代过程的策略选择,如 EPSDE^[12]、SaDE^[13]等。通过对上述文献分析不

到稿日期:2014-11-26 返修日期:2015-03-30 本文受国家自然科学基金(61165004, 61402481),河北省青年拔尖人才支持计划(冀字[2013]),河北省自然科学基金项目(F2015403046),河北省科技支撑计划(13210331),河北省教育厅青年科学基金项目(QN20131053),石家庄经济学院博士科研启动基金项目(BQ201322),江西省教育厅青年科学基金项目(GJJ14456, GJJ14373),江西理工大学博士科研启动基金项目(JXXJBS13028)资助。

汪慎文(1979—),男,博士,副教授,主要研究方向为智能计算与机器学习等;张文生(1966—),男,博士,研究员,博士生导师,主要研究方向为模式识别与机器学习等;丁立新(1967—),男,博士,教授,博士生导师,主要研究方向为智能计算及其理论;谢承旺(1974—),男,博士,副教授,主要研究方向为智能计算;郭肇禄(1984—),男,博士,讲师,主要研究方向为智能计算与并行计算。

难发现,目前 DE 研究热点之一是策略集成,特别是基于动态知识的策略集成。然而,已有的文献都集中在变异策略的动态集成上,而很少去探讨参数的动态集成与评估。因此,本文沿着这一思想去探讨参数自适应选择策略并将其应用于 DE 中。

本文第 2 节介绍差分进化算法中的参数自适应策略,是本文的核心部分;第 3 节是实验研究;最后总结全文。

2 参数自适应选择策略

本文探讨了参数自适应选择策略,其基本思想是在进化的过程中,动态评估每一个被选中的参数组合的性能,并用于指导下一个迭代过程的参数配置。因此,参数自适应策略主要涉及到参数库的建立、参数评分机制和参数配置机制 3 个方面。

2.1 参数库的建立

若缩放因子 F 值较大,扰动量也大,会引起搜索步长在一个很大的范围内波动,从而提高种群多样性,但是会降低局部搜索性能;若缩放因子 F 值较小,扰动量也小,使得新解与基准个体变化不大,算法在邻域搜索。

若杂交概率 Cr 值较小,可能只有少数几维来自变异向量,适合求解可分离的优化问题;若 Cr 值较大,使得新解与原个体相差较大,则种群多样性会提高。

总之,不同缩放因子和杂交概率展示了相互不同的特性,因此,本文期望参数在进化的不同阶段能够相互补充,以便快速求解不同类型的问题。综合考虑,本文选择尽可能覆盖缩放因子和杂交概率的赋值,即缩放因子集合 $FSet$ 和杂交概率集合 $CrSet$ 的取值均从 0.1 开始按步长为 0.1 逐步递增至 1.0,集合规模为 10。

2.2 参数的评分方法

在进化的过程中,为了动态评估参数的性能,需要记录如下信息:

$ns_{i,t}$:变异杂交策略与第 $i(i=1, \dots, 100)$ 组参数组合在第 t 代产生更优的后代的个体数量。

$nf_{i,t}$:变异杂交策略与第 $i(i=1, \dots, 100)$ 组参数组合在第 t 代产生更差的后代的个体数量。

$p_{i,t}$:第 i 组参数组合在第 t 代的得分值。在初始化的第一代,每一组参数组合采用相同的概率(即 $1/100$)选择 100 组控制参数设置。当进化代数 $t > 1$ 时,采用如下公式计算第 i 组参数组合的选择概率:

$$s_{i,t} = p_{i,t-1} * \gamma + ns_{i,t} / (ns_{i,t} + nf_{i,t} + \epsilon) \quad (1)$$

$$p_{i,t} = \frac{s_{i,t}}{\sum_{i=1}^{100} s_{i,t}} \quad (2)$$

表 1 预定义精度 VTR 和最大评价次数 MaxFES

f	f1	f2	f3	f4	f5	f6	f7	f8	f9	f10	f11	f12	f13
MaxFES	2E5	2E5	5E5	5E5	5E5	1.5E5	3E5	3E5	3E5	1.5E5	1.5E5	1.5E5	1.5E5
VTR	1E-50	1E-50	1E-8	1E-10	1E-8	1E-8	1E-3	1E-8	1E-10	1E-10	1E-20	1E-8	1E-8

为了测试算法设计的有效性,本文选择几种模型进行比较,分别为:(1)传统地应用 DE/rand/1 变异策略的 DE 算法;(2)从参数库中随机选择参数组合(Parameter Random Selection, PRS)的 DE-PRS 算法;(3)从参数库中依据参数评估结果应用随机遍历抽样法的 DE-PSS;(4)变异策略动态选择的 SaDEb。本文使用适应度评价次数 (FEs)^[15]、成功运行次数

其中, γ 为折旧率, ϵ 的作用是避免在第 t 次进化中第 i 组参数未选中导致分母为 0 的情况,本文设置 ϵ 值为 0.001。

2.3 参数的配置方法

在众多的配置方法中,轮盘赌选择法、随机遍历抽样法等是广为使用的方法。其中,随机遍历抽样方法因只需一个随机数即可为全部个体配置参数组合而被本文采用。

2.4 应用参数自适应选择策略的 DE-PSS 算法

借鉴传统的 DE 算法框架,应用参数自适应选择策略(Parameter Self-Selection, PSS)的差分进化算法(DE-PSS)的伪代码如下。

算法 1 应用参数自适应选择策略的 DE-PSS 算法

1. 设置 NP、D、 γ 、MaxFES、VTR
2. 使用公式 $x_{i,j}(0) = L_j + \text{rand}_{i,j}[0, 1](U_j - L_j)$ 初始化种群个体并评估个体,令 $FES = NP$
3. while $FES < \text{MaxFES}$ abs(Min(fit)) > VTR
4. 配置:为每一个个体使用随机遍历抽样方法分配一个参数组合
5. 变异:对种群个体用如下公式进行突变
$$V_i(t) = x_{r1}(t) + F(x_{r2}(t) - x_{r3}(t))$$
6. 修补:对越界个体用如下公式处理
$$v_{i,j} = \begin{cases} \min\{U_j, 2L_j - v_{i,j}(t)\}, & \text{if } v_{i,j}(t) < L_j \\ \max\{L_j, 2U_j - v_{i,j}(t)\}, & \text{if } v_{i,j}(t) > U_j \end{cases}$$
7. 杂交:对种群个体用如下公式进行杂交
$$\mu_{i,j} = \begin{cases} v_{i,j}(t), & \text{if } \text{rand}_j(0, 1) < Cr \text{ or } j = j_{\text{rand}} \\ v_{i,j}(t), & \text{otherwise} \end{cases}$$
8. 选择:对个体用如下公式逐一竞争选择
$$X_i(t+1) = \begin{cases} X_i(t), & \text{if } f(X_i(t)) \leq f(U_i(t)) \\ U_i(t), & \text{if } f(U_i(t)) < f(X_i(t)) \end{cases}$$
9. 评分:参数组合用式(1)和式(2)评估
10. end while

3 实验研究

本文选择 13 个有代表性的经典基准函数来测试算法的有效性,关于 13 个 Benchmark 函数的详细描述,参考文献[14]。

所有实验均在操作系统为 Windows 7(64Bit),双核 2.40 GHz 的 Intel 处理器和 2GB 内存,Matlab R2013a 的平台上完成。

实验中各算法所用到的参数均相同,分别设置如下:群体大小 30,个体维数 30,折旧率 0.8,算法运行次数为 $RunTimes = 50$,当适应度评价次数达到最大评价次数(MaxFES)或者得到了算法预定义精度(Value-to-Reach, VTR)时算法终止,最大评价次数 MaxFES 和求解精度 VTR 见表 1。

(SR)^[15] 和加速比(AR)^[15] 3 个指标来考察算法的收敛速度和算法的健壮性。

3.1 4 种算法收敛速度的比较

表 2 给出了 DE、DE-PRS、DE-PSS、SaDE 在 13 个测试函数独立运行 50 次后得到的实验统计结果。其中,“Mean”和“Std Dev”分别表示算法终止时适应度评价次数的平均值和

标准值,“NA”表示算法在 MaxFES 内均未成功收敛到预定义的精度,“AVER”表示独立运行 50 次,算法终止时所有函数上耗费的适应度评价次数的平均值。从表 2 中的“AVER”值可以看到,在相同的情况下,DE-PSS 所需要的代价最小,只需要 17.9 万次,而 DE-PSS 需要 19.2 万次, SaDE 需要 19.6 万次,DE 需要 20.9 万次,说明 DE-PSS 算法总体性能好于其他 3 个。从表中的细节数据也可以看到,DE 在 5 个函数上运行 50 次都不能收敛,DE-PRS 和 SaDE 在 2 个函数上运行 50

次都不能成功收敛,相比之下,DE-PSS 在所有的函数上都能收敛。因此,DE-PSS 更有普适性。从表 2 中还可以看出,DE-PSS 在函数 f4、f11、f12、f13 上性能更加卓越,所需的代价在 4 个算法中是最小的,而其中 f11、f12、f13 属于多模函数,这说明在面对复杂的情况时,参数自适应选择策略能够根据优化问题当前的情况自适应地做出有利于解决问题的选择,这也符合当初策略设计的初衷。

表 2 DE-PSS 与 DE、DE-PRS、SaDE 的适应度评价次数统计结果

Fun.	DE	DE-PRS	SaDE	DE-PSS
	Mean±Std Dev	Mean±Std Dev	Mean±Std Dev	Mean±Std Dev
f1	NA±NA	9.47E+04±1.42E+03	9.80E+04±3.05E+03	1.06E+05±2.46E+04
f2	NA±NA	1.54E+05±1.09E+03	1.38E+05±5.15E+03	1.62E+05±2.55E+04
f3	NA±NA	2.93E+05±1.51E+04	3.69E+05±3.27E+04	4.60E+05±6.10E+04
f4	4.53E+05±6.43E+03	NA±NA	4.21E+05±8.13E+04	3.22E+05±1.53E+05
f5	NA±NA	NA±NA	NA±NA	5.00E+05±2.42E+03
f6	1.92E+04±6.59E+02	8.35E+03±1.39E+03	1.04E+04±4.12E+03	9.15E+03±1.78E+03
f7	NA±NA	3.00E+05±1.79E+03	NA±NA	2.99E+05±1.04E+04
f8	6.44E+04±3.40E+04	1.65E+05±1.36E+05	8.93E+04±1.06E+05	9.09E+04±1.12E+05
f9	1.79E+05±3.26E+04	2.34E+05±1.02E+05	2.00E+05±1.24E+05	2.06E+05±1.07E+05
f10	9.70E+05±1.16E+03	4.09E+04±1.02E+03	1.27E+05±4.47E+04	4.41E+04±8.83E+03
f11	8.98E+04±2.60E+03	4.74E+04±3.46E+04	9.53E+04±5.75E+04	4.13E+04±1.72E+04
f12	5.12E+04±1.27E+03	2.60E+04±2.56E+04	6.78E+04±6.24E+04	2.47E+04±1.85E+04
f13	6.37E+04±1.54E+03	1.30E+05±4.58E+04	1.28E+05±4.77E+04	6.27E+04±5.53E+04
AVER	2.09E+05±6.18+0.3	1.92E+05±2.81E+04	1.96E+05±4.37E+04	1.79E+05±4.60E+04

3.2 4 种算法成功次数和加速比比较

表 3 给出了 DE-PSS 和 DE、DE-PRS、SaDE 在 13 个测试函数上成功收敛的运行次数 SR 和加速比 AR 的统计结果,其中,“(4)V(1)”表示 DE-PSS 算法“VS”DE 算法,“(4)V(2)”表示 DE-PSS 算法“VS”DE-PRS 算法,“(4)V(3)”表示 DE-PSS“VS”SaDE 算法,“NA”表示被比较算法在 50 次独立运行中均未收敛到预定义的精度(VTR)。

表 3 DE-PSS 与 DE、DE-PRS、SaDE 的成功运行次数和加速比的统计结果

Fun.	DE	DE-PRS	SaDE	DE-PSS	(4)v(1)	(4)v(2)	(4)v(3)
	SR	SR	SR	SR	AR	AR	AR
f1	0	50	50	50	NA	1.00	1.00
f2	0	50	50	41	NA	0.82	0.82
f3	0	50	50	20	NA	0.40	0.40
f4	50	0	32	30	0.60	NA	0.94
f5	0	0	0	2	NA	NA	NA
f6	50	50	50	50	1.00	1.00	1.00
f7	0	1	0	1	NA	1.00	NA
f8	49	25	40	39	0.80	1.56	0.98
f9	47	15	20	23	0.49	1.53	1.15
f10	50	50	11	50	1.00	1.00	4.55
f11	50	45	24	49	0.98	1.09	2.04
f12	50	48	32	49	0.98	1.02	1.53
f13	50	8	9	36	0.72	4.50	4.00
AVER	30.46	30.15	28.31	33.85			

从表 3 的“AVER”值可以看到,DE-PSS 成功运行次数的平均值为 33.85,高于 DE 的 30.46、DE-PRS 的 30.15、SaDE 的 28.31,这说明 DE-PSS 具有较强的健壮性,算法能根据实际优化问题自适应选择参数。

结合表 2 和表 3 可以看到,尽管 DE-PRS 的适应度评价次数平均值低于 DE,但是其成功率小于 DE,说明基于参数组合随机配置策略有助于提高解决问题的效率,但是也带来了算法的不稳定性。应用参数评估结果指导参数配置的 DE-

PSS 算法与 DE-PRS 算法相比,不仅评价次数平均值降低,而且成功率显著提高,说明本文提出的参数自适应机制不仅提高了解决问题的效率,也带来了更高的可靠性。

结束语 针对差分进化算法的参数选择的困难,以及参数组合动态配置的不足,本文着重研究参数自适应选择机制,并将其应用于差分进化算法中。通过设置参数库、设计参数评分方法和参数动态配置方法,在进化过程中,根据优化问题的特性和参数的特性,动态评估参数的性能,并用于指导下一步参数的配置。通过在 13 个测试函数上的运算比较,基于参数自适应选择策略的差分进化算法具有更高的收敛速度和鲁棒性。

参 考 文 献

- [1] Storn R, Price K. Differential evolution: a simple and efficient adaptive scheme for global optimization over continuous spaces: TR-95-012[R]. ICSI, USA, 1995
- [2] 汪慎文, 丁立新, 张文生, 等. 差分进化算法研究进展[J]. 武汉大学学报(理学版), 2014, 60(4): 283-292
Wang Shen-wen, Ding Li-xin, Zhang Wen-sheng, et al. Survey of Differential Evolution[J]. Journal of Wuhan University(Natural Science Edition), 2014, 60(4): 283-292
- [3] Liu J. A fuzzy adaptive differential evolution algorithm[J]. Soft Computing, 2005, 9(6): 448-462
- [4] Zhang J, Sanderson A. JADE: adaptive differential evolution with optional external archive[J]. IEEE Transactions on Evolutionary Computation, 2009, 13(5): 945-958
- [5] Brest J, Greiner S, Boskovic B, et al. Self-adapting control parameters in differential evolution: A comparative study on numerical benchmark problems[J]. IEEE Transactions on Evolutionary Computation, 2006, 10(6): 646-657
- [6] Fan L J, Lampinen J. A trigonometric mutation operator to differential evolution[J]. Journal of Global Optimization, 2003, 27

(1):105-129

- [7] Wang S W, Duan Y M, Shu W N, et al. Differential evolution with elite mutation strategy[J]. Journal of computational information systems, 2013, 9(3): 855-862
- [8] Sun J, Zhang Q, Tsang E. DE/EDA: A new evolutionary algorithm for global optimization[J]. Information Sciences, 2005, 169(3): 249-262
- [9] Rahnamayan S, Tizhoosh H, Salama M. Opposition-based differential evolution[J]. IEEE Transactions on Evolutionary Computation, 2008, 12(1): 64-79
- [10] Liu H, Cai Z, Wang Y. Hybridizing particle swarm optimization with differential evolution for constrained numerical and engineering optimization[J]. Applied Soft Computing, 2010, 10(2): 629-640
- [11] Wang Y, Cai Z X, Zhang Q F. Differential evolution with composite trial vector generation strategies and control parameters

[J]. IEEE Transactions on Evolutionary Computation, 2011, 15(1): 55-66

- [12] Mallipeddi R, Suganthan P, Pan Q, et al. Differential evolution algorithm with ensemble of parameters and mutation strategies [J]. Applied Soft Computing, 2011, 11(2): 1679-1696
- [13] Qin A, Huang V, Suganthan P. Differential evolution algorithm with strategy adaptation for global numerical optimization[J]. IEEE Transactions on Evolutionary Computation, 2009, 13(2): 398-417
- [14] Yao X, Liu Y, Liu G. Evolutionary programming made faster [J]. IEEE Transactions on Evolutionary Computation, 1999, 3(2): 82-102
- [15] Suganthan P, Hansen N, Liang J, et al. Problem definitions and evaluation criteria for the cec 2005 special session on real-parameter optimization[R]. Nanyang Technological University, 2005

(上接第 250 页)

结束语 标签预测在社会网络中具有广泛的应用前景。本文提出了一种改进的无偏节点标签预测算法。基于所有观察数据的训练目标的联合概率最大化与以这些数据为条件的单变量边缘预测值的不匹配现象,提出了一种改进的图模型训练方法。然后,通过对置信度的无偏估计,基于子图方法提出一种不包含额外标签数据的无偏算法用于模型的训练。Twitter 和 Pokec 数据集实验表明,与相关的标签预测算法相比,本文提出的算法在准确性和运行效率方面都得到了明显的提升。

参考文献

- [1] Sadilek A, Kautz H, Bigham J P. Finding your friends and following them to where you are[C]// Proceedings of the fifth ACM international conference on Web search and data mining. ACM, 2012: 723-732
- [2] 张子柯. 社会化标签系统的结构,演化和功能[J]. 上海理工大学学报, 2011, 33(5): 445-451
Zhang Zi-ke. Social tagging systems: structure, dynamics and function[J]. Journal of University of Shanghai For Science and Technology, 2011, 33(5): 445-451
- [3] Weinsberg U, Bhagat S, Ioannidis S, et al. BlurMe: Inferring and obfuscating user gender based on ratings[C]// Proceedings of the Sixth ACM Conference on Recommender Systems. ACM, 2012: 195-202
- [4] Hong L, Ahmed A, Gurumurthy S, et al. Discovering geographical topics in the twitter stream[C]// Proceedings of the 21st International Conference on World Wide Web. ACM, 2012: 769-778
- [5] Mahmud J, Nichols J, Drews C. Where Is This Tweet From? Inferring Home Locations of Twitter Users[C]// ICWSM. 2012: 511-514
- [6] 魏建良, 朱庆华. 基于社会化标注的个性化推荐研究进展 [J]. 情报学报, 2010, 29(4): 625-633
Wei Jian-liang, Zhu Qing-hua. Advances in Personalized Information Recommendation Based on Social Tagging[J]. Journal of The China Society for Scientific and Technical Information, 2010, 29(4): 625-633

- [7] 吴超, 周波. 基于复杂网络的社会化标签分析[J]. 浙江大学学报(工学版), 2010, 44(11): 2194-2197
Wu Chao, Zhou Bo. Complex network analysis of tag as a social network[J]. Journal of Zhejiang University (Engineering Science), 2010, 44(11): 2194-2197
- [8] Sen P, Namata G, Bilgic M, et al. Collective classification in network data[J]. AI magazine, 2008, 29(3): 93-106
- [9] Azran A. The rendezvous algorithm: Multiclass semi-supervised learning with markov random walks[C]// Proceedings of the 24th International Conference on Machine Learning. ACM, 2007: 49-56
- [10] 廖志芳, 李玲, 刘丽敏, 等. 三部图张量分解标签推荐算法[J]. 计算机学报, 2012, 35(12): 2625-2632
Liao Zhi-fang, Li Ling, Liu Li-min, et al. A Tripartite Decomposition of Tensor for Social Tagging[J]. Chinese Journal of Computers, 2012, 35(12): 2625-2632
- [11] Talukdar P P, Reisinger J, Paşca M, et al. Weakly-supervised acquisition of labeled class instances using graph random walks [C]// Proceedings of the Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2008: 582-590
- [12] 袁柳, 张龙波. 基于概率主题模型的标签预测[J]. 计算机科学, 2011, 38(7): 175-180
Yuan Liu, Zhang Long-bo. Social Tag Predication Based on Probabilistic Topic Model[J]. Computer Science, 2011, 38(7): 175-180
- [13] Bhagat S, Cormode G, Muthukrishnan S. Node classification in social networks[M]// Social network data analytics. Springer US, 2011: 115-148
- [14] Chaudhari G. High confidence predictions in social networks [D]. IIT Bombay, 2013
- [15] Java A, Song X, Finin T, et al. Why we twitter: understanding microblogging usage and communities[C]// Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 Workshop on Web Mining and Social Network Analysis. ACM, 2007: 56-65
- [16] Bae S H, Halperin D, West J, et al. Scalable Flow-Based Community Detection for Large-Scale Network Analysis[C]// 2013 IEEE 13th International Conference on Data Mining Workshops (ICDMW). IEEE, 2013: 303-310