

基于遗传编程的中国股票市场有效性检验

王红霞¹ 曹 波²

(北京青年政治学院计算机系 北京 100102)¹ (北京工业大学计算机学院 北京 100124)²

摘 要 现代资本市场理论与金融投资实践之间存在着有效市场假说与技术分析之间的矛盾,使用流行的技术交易规则检验股票市场有效性可能导致两种结论偏差。遗传编程使用树形结构表示问题的候选解,可以很好地描述技术交易规则。利用遗传编程算法生成一种技术交易策略,并用其检验上证综合指数和 5 个沪深股市个股。回测结果表明,提出的方法相对于“买入_持有”策略能够获得超额收益,并且优于常用的流行技术指标,也说明我国股票市场并未达到弱式有效。

关键词 遗传编程,市场有效性,中国股票市场

中图法分类号 TP391 文献标识码 A

Chinese Stock Market Efficiency Testing Based on Genetic Programming

WANG Hong-xia¹ CAO Bo²

(Computer Department, Beijing Youth Politics College, Beijing 100102, China)¹

(School of Computer Science, Beijing University of Technology, Beijing 100124, China)²

Abstract There is a contradiction between the modern capital market theory and the financial investment practice. And the contradiction is about the effective market hypothesis and the technical analysis. The use of the popular technology trading rules to examine the effectiveness of the stock market may lead to two types of conclusion deviation. The tree structure is used to represent the candidate solutions in genetic programming which can well describe the technical trading rules. The genetic programming algorithm is used to generate technical trading strategy in this paper. The strategy is used to test Shanghai indexes and five stocks in the Shanghai and Shenzhen stock markets. The back test results show that genetic programming generates the best technical trading strategy with significant excess profit compared with buy-and-hold strategy and the usual popular technical indicator. Therefore, the conclusion can be made that Chinese stock market has not achieved weak-form efficiency.

Keywords Genetic programming, Market efficiency, Chinese stock market

1 引言

现代资本市场理论与金融投资实践之间的重大分歧之一是有有效市场假说与技术分析之间的矛盾。技术分析认为可以通过分析价格、成交量等历史交易数据预测股票的走势,从而指导交易并获取超额收益,而有效市场假说则认为在弱式有效的市场中,当前股票价格已经充分反映了历史交易信息,技术分析不可能获得超额收益。由于有效市场假说与技术分析存在对立,因此可以通过检验技术分析的有效性间接检验市场的有效性。

在对中国股票市场的技术分析有效性实证检验中,学者们常用比较分析法,与“买入_持有”策略相比,流行的技术指标生成的交易策略常能获得超额收益。魏玉根实证检验移动平均线指标在 1992 年 11 月—1999 年 10 月的上证综指和深证综指均可获得稳定的超额收益^[1];邓杰等在不考虑交易成

本的情况下,运用移动平均线和“阻力线_支撑线”等技术交易规则,实证研究 1990 年 12 月—2008 年 3 月上证综指的有效性,发现仅“阻力线_支撑线”规则可以获得超额收益^[2];林赵华采用 CCI(Commodity Channel Index)指标对 1991 年 5 月—2005 年 11 月的上证综指和深证综指进行检验,发现用 CCI 指标能够产生超额收益^[3];曾劲松以 1998 年 1 月—2004 年 8 月间的上证综指、深证综指以及部分个股为研究对象,发现移动平均线规则不能持续获得超额收益^[4]。流行的技术指标是在多年市场起伏中由投资者逐步选择出来的,实质上这些检验和指标的筛选均使用同一段历史数据,缺乏对技术交易规则有效性的样本外测试,仍存在“数据探测”问题。因此,流行技术指标生成的交易策略战胜市场并不能完全否定中国股票市场的有效性。

针对这一问题,基于遗传编程算法构建了中国股票市场的技术交易系统。该系统以历史交易数据(训练集)作为系统

本文受北京社科联 2015 青年社科人才资助项目(2015SKL029),北京青年政治学院 2016 年委托课题(WT201606),北京市教委科研计划项目(SM201411626006)资助。

王红霞(1979—),女,博士,副教授,主要研究方向为智能信息处理、软件与信息服务、数据挖掘等,E-mail:whx4617@163.com;曹波(1980—),男,博士生,主要研究方向为机器学习、计算经济、量化交易分析等。

的输入,通过遗传编程算法进化生成最优交易策略,并在测试集中检验最优交易策略的盈利能力,从而间接检验市场有效性假说。该系统采用移动检验的方法,在测试集中每次检验完最优技术交易策略后,测试集中的部分数据将转变成训练集数据。由于在测试集中检验的最优交易策略数据是由遗传编程算法在训练集中随机生成的,避免了“数据探测”问题,因此对该交易系统性能的评价是对中国股票市场有效性的更可靠检验。

2 遗传编程算法

遗传编程(Genetic Programming, GP)算法^[5]是由美国斯坦福大学 Koza 教授提出的,是利用达尔文生物进化思想设计的一种优化算法。遗传编程算法在生成个体树时,从终端集中随机选择变量构建叶子节点,从函数集中随机选择函数构建非叶子节点,随机生成由个体树构成的初始种群。初始种群依据优胜劣汰的进化原则,经过复制、交叉、变异等遗传操作,在适应度函数的指引下逐渐进化出求解问题的最优个体树,即最优解。

2.1 函数集和终端集

遗传编程算法通过个体树表达问题的可行解,生成个体树时,需设定函数集和终端集。函数集表示对数值如何处理,而终端集则表示输入变量。

函数集 F 包含 N 个函数,即:

$$F = \{f_1, f_2, \dots, f_i, \dots, f_N\} \quad (1)$$

其中, $1 \leq i \leq N$ 。

函数集内的函数 f_i 可以是算术运算符(如+, -, ×, ÷等)、标准数学函数(如 sin, cos, exp, log 等)、布尔运算符(如 and, or, not 等)。

终端集 T 包含 M 个变量,即:

$$T = \{t_1, t_2, \dots, t_i, \dots, t_M\} \quad (2)$$

其中, $1 \leq i \leq M$ 。

终端集中的终止符可以是一些随机生成的常量,也可以是一些表示问题输入数据的变量。

遗传编程算法通过随机从函数集和终端集中选择一定量的函数和终止符生成个体树。如图 1 所示,该个体树表示函数表达式 $\max(x+x, x+3*y)$, 其中圆圈部分称为函数(max, +, *), 从函数集中选择生成; 叶节点上的变量以及常数(X, Y, 3)称为终端, 从终端集中选择生成。

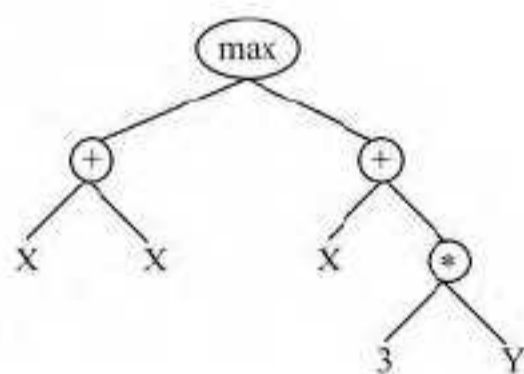


图 1 函数表达式 $\max(x+x, x+3*y)$ 的个体树

2.2 遗传算子

遗传编程算法主要有复制、交叉以及变异 3 个遗传算子,正是由于遗传算子的作用,个体树的结构才能发生变化,从而生成新的个体。

2.2.1 复制算子

复制算子是指按照一定规则把父代种群中适应度较高的个体复制遗传到子代种群中。一般而言,个体的适应度越高,

被选择复制的机会就越大。选择方法主要有比例选择法、分级选择法、竞技选择法以及截取选择法,其中竞技选择法是最常用的选择方法。

2.2.2 交叉算子

交叉算子是指父代两个个体相互交换部分信息。交叉时,从父代种群中选择两个个体,每个父代个体随机选择一个交换点,于是便产生一个以交换点为根的子树,该子树包括交换点以下的所有子树,此子树称为交换段。将第一个父代个体的交换段删除后,再把第二个父代个体的交换段插入到交换点处,这样就产生了第一个子代个体,同样操作可产生第二个子代个体。如图 2 所示。

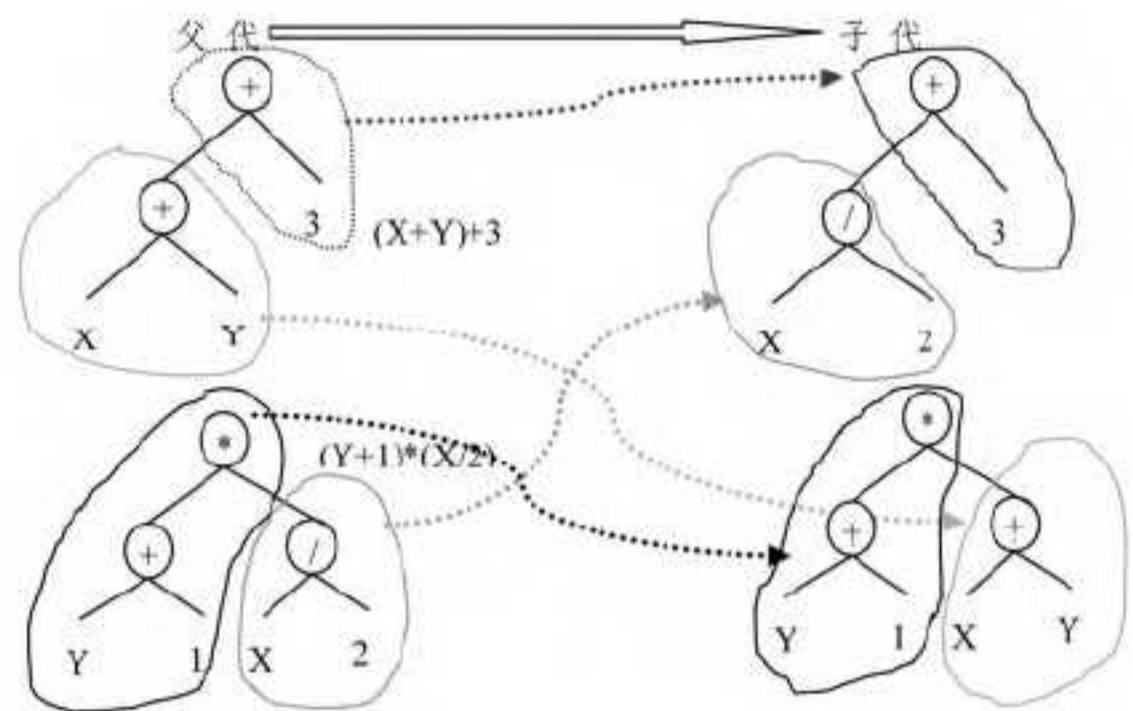


图 2 个体树 $(X+Y)+3$ 和个体树 $(Y+1)*(X/2)$ 的交叉过程

2.2.3 变异算子

变异算子是指改变个体的某部分节点。变异时,从父代种群的个体中随机选取某个节点作为变异点,并用新生成的个体树代替该变异点以下的子树。如图 3 所示。

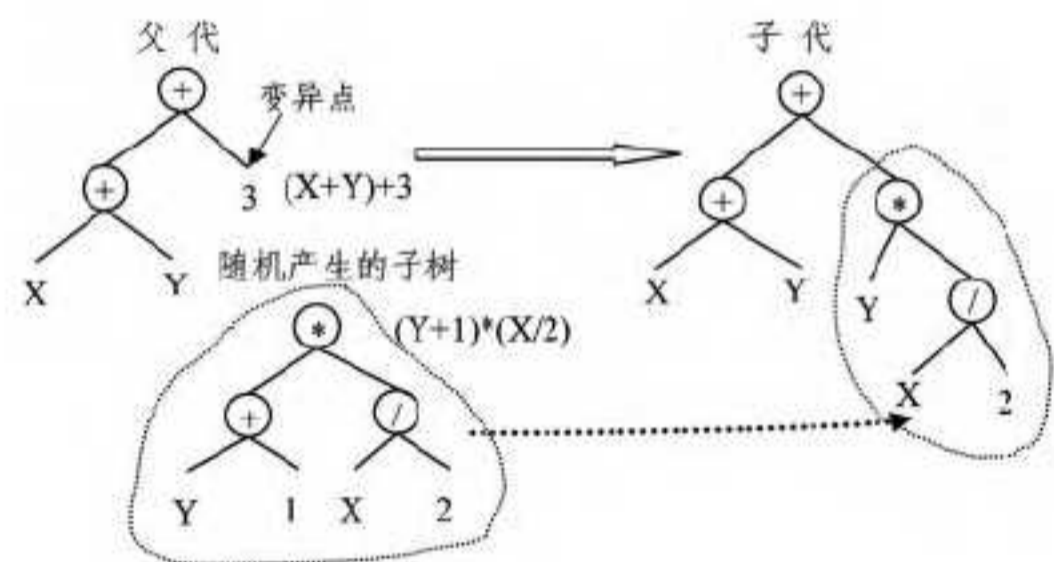


图 3 个体树 $(X+Y)+3$ 中节点“3”的变异过程

2.3 适应度函数

个体生成之后,需要知道该个体对环境的适应程度,适应度高的存活下来的可能性也更高,或者用于进行交叉或变异遗传操作形成新个体,从而将优良基因遗传给下一代。对于不同的问题,可以设置不同的适应度函数,例如,对于拟合问题,可以将拟合均方误差为适应度;对于交易策略问题,则可以将收益率的最大回撤作为适应度。

2.4 遗传编程算法流程

遗传编程算法首先随机生成初始种群,然后进入循环迭代过程。在循环迭代过程中,算法首先计算个体的适应度值,并判断是否满足迭代终止条件,若符合则迭代终止,否则按照一定选择方法从父代种群中随机选择个体进行复制、交叉和变异等遗传操作并生成新个体。所有的新个体构成下一代新种群,新种群再进行适应度值计算、迭代终止条件判定以及通过遗传算子生成新个体的循环过程。遗传编程算法流程图如图 4 所示。

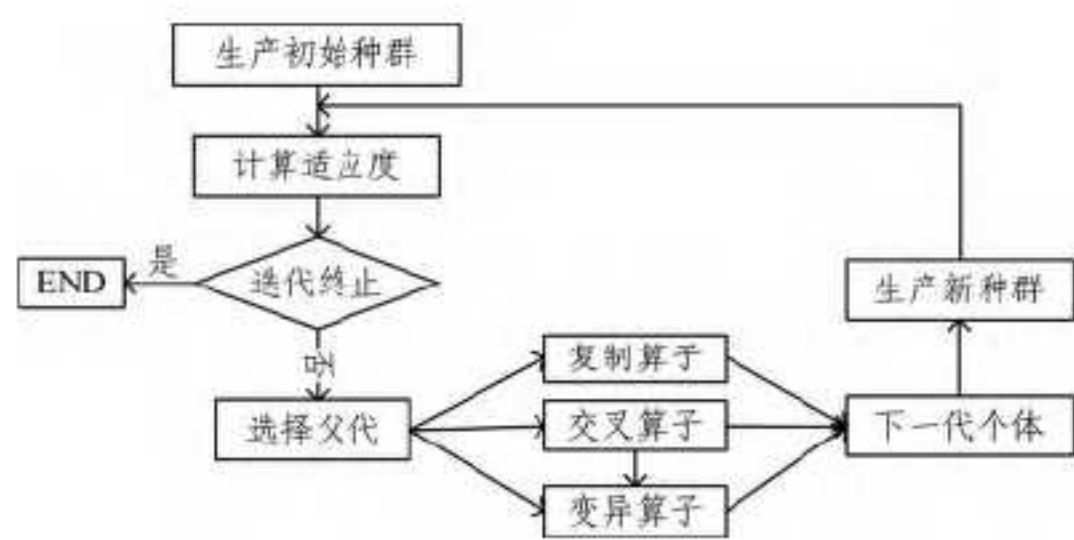


图4 遗传编程算法流程图

3 基于遗传编程算法的交易系统

遗传编程算法作为一种较为通用的优化方法，根据不同的优化对象和优化目标，在终端集、函数集、适应度函数、进化规则等重要控制参数的设计上具有很大的灵活性。本研究将使用遗传编程算法搜索中国股票市场的最优技术交易策略，因此结合问题的具体特点进行如下设计。

3.1 终端集的设置

技术交易策略通过分析历史价格和历史成交量来做出买卖决策。为了使算法进化出技术交易策略，该交易系统设置的终端集包括历史交易数据、技术指标数据、布尔型常数和实型常数，其中技术指标数据由历史交易数据衍生而出来，如表1所列。

表1 交易系统的终端集

数据类型	数据
历史交易数据	开盘价
	最高价
	最低价
	收盘价
	成交量
技术指标数据	双移动平均线交叉指标
	相对强弱指标
	威廉指标
	顺势指标
	快速随机指标
	慢速随机指标
	平滑异同移动平均线指标
布尔型数据	逻辑真
	逻辑假
实型数据	(0,1)之间均匀分布的随机数

3.2 函数集的设置

设定的函数集包括算术运算符、比较运算符和逻辑运算符等，如表2所列。

表2 交易系统的函数集

函数类型	函数名称	返回值
算术运算符	+, -, *, /	序列实数
比较运算符	>, <	布尔型
逻辑运算符	and, or	布尔型

3.3 适应度函数的选取

适应度函数的选取对一个交易系统的成败起着至关重要的作用。在投资实践中，往往采用最大回撤评估技术交易策略的性能。最大回撤是指技术交易策略在选定周期内任一历史时点往后推，资金净值走到最低点时的收益率回撤幅度的最大值。最大回撤用来描述策略可能出现的最差的情况。记 CR_i 是到第 i 天的累计收益，截止到第 k 天的最大回撤为：

$$\max Drawdown_k = \min(CR_k - \max_{1 \leq i \leq k} CR_i) \quad (3)$$

交易系统进化的技术交易策略以最大回撤作为评估标准。技术交易策略的最大回撤越小，则适应度越高。

3.4 进化规则

遗传编程算法作为一种优化方法，为求解复杂问题提供了通用的框架。系统利用遗传编程算法生成技术交易策略时，首先生成初始种群，然后利用复制、交叉和变异等遗传算子进化出最大回撤较小的技术交易策略。在利用遗传编程算法对生成技术交易策略建模时，根据求解问题设定算法的进化参数，以便能进化出技术交易策略。具体设置的进化参数如表3所列。

表3 交易系统的参数设置

参数	取值
进化代数	100
群体规模	500
锦标赛规模	2
复制比率	0.2
交叉比率	0.6
变异比率	0.2

3.5 基于遗传编程算法的技术交易系统回测流程

基于遗传编程算法的技术交易系统采用移动检验的方式，每次检验分为训练阶段和测试阶段。在训练阶段，利用遗传编程算法从历史数据（训练集）中进化搜索技术交易策略，进化完毕后输出最优技术交易策略，在测试阶段，最优技术交易策略根据未来数据（测试集）产生交易信号（0即卖出，1即买入），并根据交易结果计算收益率。一次交易过程（开仓和平仓）完成后，将用于测试最优策略的未来数据放入历史数据中，并开始新一轮的训练和测试过程。技术交易系统的回测流程如图5所示。



图5 基于遗传编程算法的技术交易系统回测流程

4 实验与结论

选取上证综合指数和部分个股作为研究对象。为了验证所提方法的有效性，在回测期间内比较了遗传编程算法生成的技术交易策略、利用流行技术指标生成的技术交易策略、以及“买入持有”策略的盈利能力。表4列出了流行技术指标以及所用到的参数。

表4 流行技术指标以及所用到的参数

指标名称	英文缩写	参数
双移动平均线交叉	DMAC	6, 24
相对强弱指标	RSI	12, [30, 70]
威廉指标	WPR	14, -50
顺势指标	CCI	24, 0
快速随机指标	FSTO	24, 6, [20, 60]
慢速随机指标	SSTO	24, 6, [20, 60]
平滑异同移动平均线	MACD	12, 26, 9

实验使用的数据包括：上证综指在2010年4月20日—2014年12月31日期间日K线历史数据，以及相同时段的5个沪深股市个股日K线历史数据。上证综指的交易成本为

0.04%, 个股的交易成本为 0.26%。表 5 列出了该系统生成的遗传编程(GP)最优技术交易策略在相应测试期的收益率、

“买入_持有”策略(B&H)收益率以及表 4 所列的流行技术分析指标收益率的统计结果。

表 5 测试品种的遗传编程策略 GP、“买入_持有”策略 B&H、流行技术指标的收益率统计结果(%)

测试品种	GP	B&H	DMAC	RSI	WPR	CCI	FSTO	SSTO	MACD
上证综指	13.6	8.5	-8.2	3.4	-6.6	-4.0	8.6	8.0	-1.3
000002	58.2	36.0	3.2	-35.3	-21.3	15.6	9.2	13.5	-11.2
000623	17.5	-36.0	40.4	27.5	13.5	-38.6	7.7	36.2	10.4
600000	17.6	-17.9	3.0	1.3	-2.6	-0.5	2.7	-3.0	-13.6
600019	9.6	-22.4	4.0	2.8	-1.5	-19.2	-1.6	3.3	-9.5
601318	20.5	-8.2	12.2	14.2	-42.3	-6.7	-14.3	-2.3	18.5
平均值	22.8	-6.7	9.1	2.3	-10.1	-8.9	2.1	9.3	-1.1

从表 5 统计结果可以看出,在所有测试品种中,遗传编程算法(GP)生成的技术交易策略均获得了正收益,平均收益率高达 22.8%,远优于其他策略的收益率;流行技术指标中,DMAC 和 SSTO 收益较好,在测试期内,各测试品种的平均收益率分别达到 9.1%和 9.3%,所有流行技术指标的收益率波动较大,遗传编程算法生成的技术交易策略与“买入_持有”策略相比较,获得了 29.5%的超额收益。以上结果表明了基于遗传编程算法生成的技术交易策略的有效性,从而说明了我国股票市场并未达到弱式有效。

结束语 尽管影响股票市场的因素很多^[6],宏观因素包括政策法规、利率调整、国际金融市场的变化、重要人物的讲话等,微观因素包括行业发展、区域经济、公司发展等,但市场的有效性与投资者的利益直接相关,并深远地影响着资本资产配置和定价的效率^[7,8]。长期以来,技术交易规则的有效性与有效市场假说之间的关系一直是学术界的讨论热点。因为可以通过检验技术分析的有效性间接检验股票市场的有效性,所以文中构建了基于遗传编程算法的中国股票市场的有效性检验系统。该系统以历史交易数据(训练集)作为系统的输入,通过遗传编程算法进化生成最优交易策略,并在测试集中检验最优交易策略的盈利能力,从而间接检验市场的有效性。系统采用移动检验的方法,在测试集中每次检验完最优技术交易策略后,测试集中的部分数据将转变成训练集数据。

最后利用该算法对上证综合指数和 5 个沪深股市个股进行实证检验,回测结果表明,提出的方法相对于“买入_持有”策略能够获得超额收益,并且优于常用的流行技术指标,这说明我国股票市场并未达到弱式有效。

参 考 文 献

- [1] 魏玉根. 技术交易系统与我国股票市场有效性的实证分析[J]. 经济科学, 2000(2): 56-63
- [2] 邓杰, 唐国兴. 中国股票市场技术交易规则有效性的实证研究[J]. 华东经济管理, 2009, 23(5): 135-140
- [3] 林赵华. 股票投资周 CCI 指标策略收益的检验分析[J]. 广西财经学院学报, 2009, 22(2): 64-69
- [4] 曾劲松. 技术分析与中国股票市场有效性[J]. 财经问题研究, 2005(8): 27-30
- [5] Koza J R. Genetic Programming: On the Programming of Computers by Means of Natural Selection [M]. Cambridge: The MIT Press, 1992
- [6] 朱振国, 宋军, 乜堪雄. 基于 Vague 集相似度量的股票选择[J]. 计算机科学, 2008, 35(7): 199-212
- [7] 瞿慧. 基于遗传编程的上证 50 指数技术交易规则研究[J]. 管理科学, 2010, 23(5): 103-113
- [8] 瞿慧, 刘焯, 李娟. 基于遗传编程的中国股票市场有效性新检验[J]. 控制与决策, 2011, 347(23): 137-142

(上接第 515 页)

- [8] Leavitt N. Payment applications make e-commerce mobile [J]. Computer, 2010, 43(12): 19-22
- [9] Gu Rui-jun, Yao Juan, Wang Jia-cai. Research on mobile payment technology and business models in China under e-commerce environment[J]. Future Generation Information Technology, 2010, 6485: 334-343
- [10] 程斌, 雷明海. 地下停车场停车智能化方案介绍[J]. 智能建筑电气技术, 2014, 8(5): 84-86
- [11] Gressmann B, Klimek H, Turau V. Towards ubiquitous indoor location based services and indoor navigation[C]//Workshop on Positioning Navigation and Communication(WPNC), 2010: 107-112
- [12] 蔡朝晖, 夏溪, 胡波, 等. 室内信号强度指纹定位算法改进[J]. 计算机科学, 2014, 41(11): 178-181
- [13] Schulcz R, Varga G, Toth L. Indoor location services and context-sensitive applications in wireless networks[C]//2010 Inter-

- national Conference on Indoor Positioning and Indoor Navigation(IPIN), 2010: 1-10
- [14] Kushki A, Plataniotis K N, Venetsanopoulos A N. Intelligent Dynamic Radio Tracking in Indoor Wireless Local Area Networks[J]. IEEE Transactions on Mobile Computing, 2010, 9(3): 405-419
- [15] Cheng Y Y, Lin Y Y. A new received signal strength based location estimation scheme for wireless sensor network[J]. IEEE Transactions on Consumer Electronics, 2009, 55(3): 1295-1299
- [16] Wu K S, Xiao J, Yi Y W, et al. CSI-based indoor localization[J]. IEEE Transactions on Parallel and Distributed Systems, 2013, 24(7): 1300-1309
- [17] 毛勤, 曾碧, 叶林锋. 改进的室内移动机器人模糊位置指纹定位研究[J]. 计算机科学, 2015, 42(11): 170-173
- [18] 张明洋, 陈剑, 闻英友, 等. 基于滑动窗口最长公共子序列 Wi-Fi 指纹定位算法[J]. 东北大学学报(自然科学版), 2014, 35(10): 1390-1394