

北京航空航天大学计算机学院

本科生毕业设计（论文）开题报告

论文题目： 基于搜索的软件工程技术在金融优化问题中的应用研究

学生姓名： 王劭阳

学生学号： 12061163

专 业： 计算机科学与技术

指导教师： 任健

学院(系)： 计算机学院

北京航空航天大学计算机学院

2016 年 10 月 15 日

填表说明

1. 根据《北京航空航天大学关于本科生毕业设计(论文)手册》要求，每位学生必须认真撰写《本科生毕业设计（论文）开题报告》。
2. 每位学生应在指导教师的指导下认真、实事求是地填写各项内容。文字表达要明确、严谨，语句通顺，条理清晰。外来语要同时用原文和中文表达，第一次出现的缩写词，须注出全称。
3. 开题报告中引用的参考文献阅读数量不少于 10 篇，其中外文资料应不少于总篇数的 30%。参考文献的书写请参照《北京航空航天大学关于本科生毕业设计(论文)手册》。
4. 本科生毕业设计（论文）开题报告总字数不少于 3000 字。
5. 请用宋体小四号字体填写，并用 A4 纸打印，于左侧装订成册。
6. 该表填写完毕后，须请指导教师审核，并签署意见。
7. 《北京航空航天大学计算机学院本科生毕业设计（论文）开题报告》将作为答辩资格审查的主要材料之一，占 5 分。
8. 本表格不够可自行扩页。

本科生毕业设计(论文)开题报告

学生姓名	王劭阳	学号	12061163	专业班级	130616
指导教师	任健	职称	讲师	单 位	计算机学院
题目类型	<input type="checkbox"/> 理论研究 <input checked="" type="checkbox"/> 应用研究 <input type="checkbox"/> 综述 <input type="checkbox"/> 其它 如选“其它”，请自定义题目类型_____				
题目来源	<input checked="" type="checkbox"/> 校内 <input type="checkbox"/> 校外				
毕业设计 (论文) 题 目	基于搜索的软件工程技术在金融优化问题中的应用研究				

开题报告（阐述课题的目的、意义、研究现状、研究内容、研究方案、进度安排、预期结果、参考文献等）

一、研究目的和意义

金融领域内的优化问题一直以来是让众多投资人、学者、企业感兴趣的问题之一。股票涨跌问题则是金融优化问题的重要代表。

股票的涨跌可能与这些事件相关：

- 供求关系：买股票需求上涨导致股票价格上涨；
- 上市公司盈利能力：和股票相关的上市公司业绩增加导致股票价格上涨；
- 大众投资者信心：投资者们倾向于认为股票价格会涨，市场信心足，导致股票价格上涨；
- 周边市场或宏观形势影响：周边国家的股票上涨导致我们的股票随之上涨；
- 未来政策、形势等发展趋势：如果大多数投资者都认为以后一段时间风调雨顺国泰民安、经济持续健康发展，那么股票就会涨；
- 其他投资品种的收益高低：存款、其他投资收益率降低导致股票价格上涨。

本课题旨在通过研究包括但不限于这些事件与股票涨跌之间的联系，得出一套通过事件的发生情况来预测股票涨跌的模型，投资者可以据此以预测股票市场的动向，并为金融领域内的优化问题提供一种解决方案。

二、研究现状

2.1 使用基于搜索的软件工程技术预测股票涨跌的方案

国内外使用软件工程技术来预测股票涨跌的方案很多。从研究数据上来看，有通过研究股票历史数据分析的，有通过研究社交网络上的舆情来分析的，有通过研究相关事件的发生来分析的。从算法上来分析有使用梯度下降算法的，有使用遗传算法的，有使用 RBF 神经网络的，有使用支持向量机的。

2.2 遗传算法

遗传算法（Genetic Algorithm）是一类借鉴生物界的进化规律（适者生存，

优胜劣汰遗传机制）演化而来的随机化搜索方法。它是由美国的 J.Holland 教授 1975 年首先提出，其主要特点是直接对结构对象进行操作，不存在求导和函数连续性的限定；具有内在的隐并行性和更好的全局寻优能力；采用概率化的寻优方法，能自动获取和指导优化的搜索空间，自适应地调整搜索方向，不需要确定的规则。遗传算法的这些性质，已被人们广泛地应用于组合优化、机器学习、信号处理、自适应控制和人工生命等领域。它是现代有关智能计算中的关键技术。

遗传算法也是计算机科学人工智能领域中用于解决最优化的一种搜索启发式算法，是进化算法的一种。这种启发式通常用来生成有用的解决方案来优化和搜索问题。

2.3 卷积神经网络

卷积神经网络是近年发展起来，并引起广泛重视的一种高效识别方法。20 世纪 60 年代，Hubel 和 Wiesel 在研究猫脑皮层中用于局部敏感和方向选择的神经元时发现其独特的网络结构可以有效地降低反馈神经网络的复杂性，继而提出了卷积神经网络（Convolutional Neural Networks-简称 CNN）。现在，CNN 已经成为众多科学领域的研究热点之一，特别是在模式分类领域得到了更为广泛的应用。K.Fukushima 在 1980 年提出的新识别机是卷积神经网络的第一个实现网络。随后，更多的科研工作者对该网络进行了改进。其中，具有代表性的研究成果是 Alexander 和 Taylor 提出的“改进认知机”，该方法综合了各种改进方法的优点并避免了耗时的误差反向传播。

由于 CNN 的特征检测层通过训练数据进行学习，所以在使用 CNN 时，避免了显示的特征抽取，而隐式地从训练数据中进行学习；再者由于同一特征映射面上的神经元权值相同，所以网络可以并行学习，这也是卷积网络的一大优势。

2.4 使用 CUDA 加速计算

CUDA(Compute Unified Device Architecture)，是显卡厂商 NVIDIA 推出的运算平台。CUDA 是一种由 NVIDIA 推出的通用并行计算架构，该架构使 GPU 能够解决复杂的计算问题。它包含了 CUDA 指令集架构（ISA）以及 GPU 内部的并行计算引擎。CUDA 将 GPU 视为一个并行数据计算的设备，对所进行的计算进行分配和管理。在 CUDA 的架构中，这些计算不再像过去所谓的 GPGPU 架构那样必须将计算映射到图形 API（OpenGL 和 Direct 3D）中，因此对于开发者来说，CUDA 的开发门槛大大降低了。

由于 GPU 的特点是处理密集型数据和并行数据计算，因此 CUDA 非常适合需要大规模并行计算的领域。目前 CUDA 已经广泛的应用在图形动画、科学计算、地质、生物、物理模拟等领域。

三、研究内容

3.1 数据说明

研究数据包括 2014~2016 两年内的每日股市开盘价、收盘价、最高价、最低价；2014~2016 两年内的每日发生的可能相关的事件，事件通过事件类型、事件性质（积极或中立或消极）、发生时间、发生次数来定义。

3.2 使用遗传算法对数据建模的研究

遗传算法是解决搜索问题的一种通用算法，本课题首先尝试使用遗传算法对数据进行建模。以每日股市最低价为基准，设 $y \in \{0, 1\}$ 为每日股市相较于前日的涨跌情况，0 代表下跌或不变，1 代表上涨。算法通过训练集进行训练后，希望得出一个假设函数 $h_{\theta}(X) \in \{0, 1\}$ ，其中 X 代表事件向量， θ 代表训练得到的参数向量，能够相对准确的预测每日的涨跌。

遗传算法将参数向量编码作为个体的基因序列，首先随机生成一个种群，种群中包括初始的个体，然后不断迭代产生新的种群，直到种群中的某个个体可以足够好的进行预测。每次迭代产生新个体的过程包括两个步骤：交叉、变异，交叉过程（如图 3.2.2 所示）随机选出两个个体交叉产生新的个体；变异过程扫描种群内的所有个体，按一定概率对其中的基因做突变。

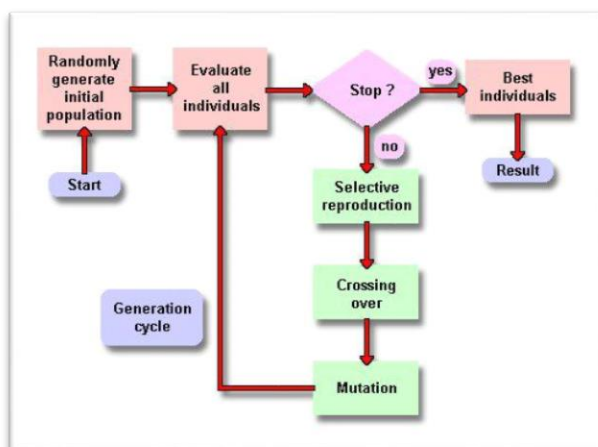


图 3.2.1 遗传算法流程

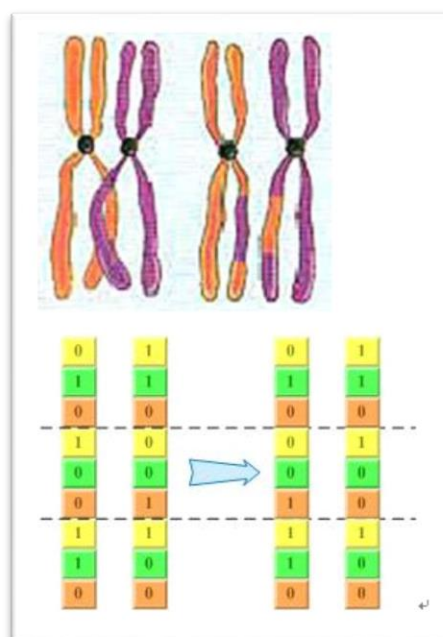


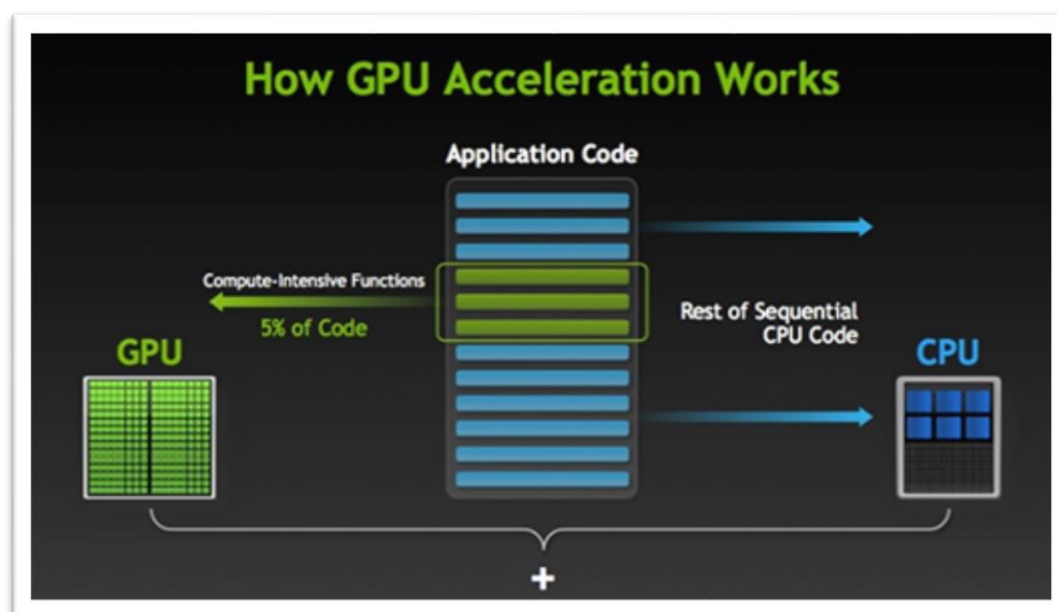
图 3.2.2 两个体交叉的过程

3.3 使用卷积神经网络对数据建模的研究

卷积神经网络（CNN）是目前较为先进的模式识别算法之一。它可以通过学习一种多层网络结构，实现复杂函数逼近，并体现它对于输入样本数据的强大的本质特征的抽取能力。在使用遗传算法得出模型后，本课题尝试使用 CNN 来对数据进行建模，尝试得出更优的模型。

3.4 使用 NVIDIA CUDA 架构优化计算的研究

遗传算法和 CNN 的执行中都需要做大量的计算，其中很多计算是可以并行化的。GPU 与 CPU 相比，核心数多很多，并行计算的能力强很多。本课题的研究之一是使用 CUDA 计算架构，将大规模的并行计算转移至 GPU 进行，来提高算法的执行效率，加快解收敛的速度。



四、研究方案和技术路线

本课题使用 python 编程实现上述的算法,在实现卷积神经网络时,使用 Caffe 框架和 Anaconda 环境辅助开发,并且使用 CUDA 计算框架进行加速,比较无 GPU 加速和有 GPU 加速两种情况的性能,然后对比两种算法的优劣。

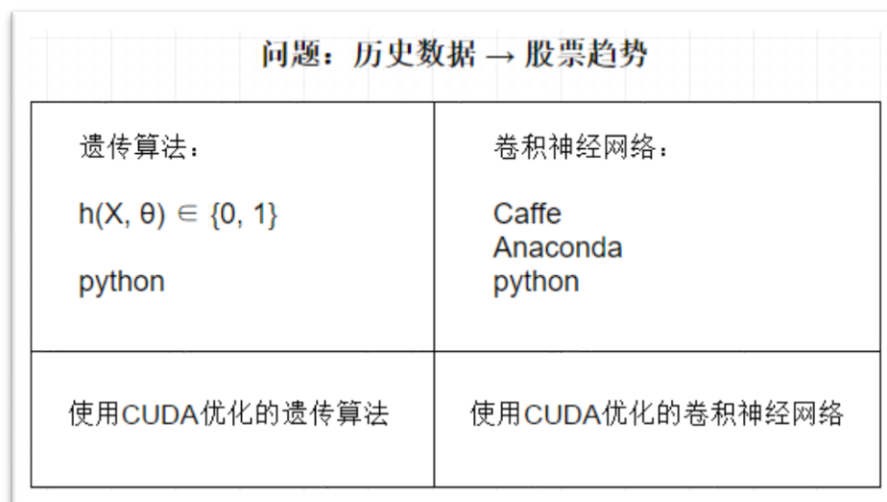


图 4.1 技术路线

五、关键技术或难点

论文研究内容中可能存在的理论问题、关键技术或技术难点。

六、预期结果

- 得到使用遗传算法对历史数据建立的模型,预测股票涨跌的成功率大于 52%
- 得到使用卷积神经网络对历史数据建立的模型,预测股票涨跌的成功率大于 52%
- 编程实现使用和不使用 GPU 加速的两种算法
- 对比使用和不使用 GPU 加速的两种算法性能
- 对比两种算法对问题建模的效果

七、进度安排

2016 年 9 月~2016 年 10 月: 学习相关技术,准备开题

2016 年 10 月~2016 年 11 月: 完成遗传算法部分的代码实现及模型训练

2016 年 11 月~2017 年 2 月: 完成卷积神经网络部分的代码实现及模型训练

2017 年 2 月~2017 年 3 月：完成 GPU 加速部分的代码并比较性能

2017 年 3 月~2017 年 4 月：继续对算法调参，比较两种算法的性能

2017 年 4 月~2017 年 5 月：完成毕设论文的撰写，准备答辩

八、参考文献

- [1] 李征, 巩敦卫, 聂长海,等. 基于搜索的软件工程研究专题前言[J]. 软件学报, 2016, 27(4):769-770.
- [2] 李峰, 韩祝华, 封二英. 基于机器学习的金融数据分析研究[J]. 科学与财富, 2016(6).
- [3] 王秀峰. 遗传算法在金融市场中的应用[C]// 中国控制会议. 1998.
- [4] 邓方舟. 基于遗传算法的投资组合在金融领域中的应用[J]. 中国科技纵横, 2015(2):41-41.
- [5] 陈为民, 马超群. 支持向量机方法及其在金融中的应用与前景[J]. 金融经济:理论版, 2006(12):114-115.
- [6] 赵志勇, 王峰, 李元香. 基于深度学习的股票市场预测[J]. 2014.
- [7] 张佳康, 陈庆奎. 基于 CUDA 技术的卷积神经网络识别算法[J]. 计算机工程, 2010, 36(15):179-181.
- [8] 张舒. GPU 高性能运算之 CUDA[M]. 中国水利水电出版社, 2009.
- [9] Harman M, Mcminn P, Souza J T D, et al. Search Based Software Engineering: Techniques, Taxonomy, Tutorial[C]// Empirical Software Engineering and Verification - International Summer Schools, LASER 2008-2010, Elba Island, Italy, Revised Tutorial Lectures. 2012:2009--2010.
- [10] Pospichal P. GPU-Based Acceleration of the Genetic Algorithm[J]. Gecco Competition, 2009.
- [11] Pospichal P, Jaros J, Schwarz J. Parallel Genetic Algorithm on the CUDA Architecture[C]// Applications of Evolutionary Computation, Evoapplicatons 2010: Evocomplex, Evogames, Evoiasp, Evointelligence, Evonum, and Evostoc, Istanbul, Turkey, April 7-9, 2010, Proceedings. 2010:442--451.
- [12] Shah D, Zhang K. Bayesian regression and Bitcoin[J]. Computer Science, 2014:409 - 414.
- [13] Turchenko V, Luczak A. Caffe: Convolutional Architecture for Fast Feature Embedding[J]. Eprint Arxiv, 2014:675-678.

指导教师意见（课题难度是否适中、工作量是否饱满、进度安排是否合理、工作条件是否具备、是否同意开题等）：

指导教师签名：_____

年 月 日

学院（系）意见：

审 查 结 果： ☐ 同 意 ☐ 不 同 意

学院（系）负责人签名：_____

年 月 日