

高频交易系统开发的各个阶段是怎样的？

[关注问题](#)[写回答](#)[邀请回答](#)[好问题](#)[添加评论](#)[分享](#) ...

1 个回答

默认排序



匿名用户

2 人赞同了该回答

高频交易系统^Q的开发大致可以分为三个阶段：研究阶段、模型阶段和实现阶段。每一个阶段都有自己的内部过程和子系统。在建立模型之前，抽象化的交易想法可能会在数学上存在些逻辑错误和缺陷。模型阶段就是要验证你的交易想法，找出逻辑错误，评估备选方案，帮助他人理解交易想法背后的数学机制^Q。模型阶段解决开发过程前期的不确定性，考察数学模型的可行性以及在不同市场环境下盈利是否稳健同样，你还可以预估一下实施阶段的工作量。模型阶段从数据准备和分类开始，经过模型编程，到最后测试和优化。

模型阶段遵照设计文档中的交易思路、数学计算和数据要求等。这一阶段强迫研发团队对数学模型清晰定义，并解决相关问题(比如数据处理)，另外，这一阶段还要定义未来软件实施的各种要求。

收集数据

清理过的历史数据是进行成功历史回测^Q和优化的基础回测本身也许会因为数据上已知和未知的错误而产生问题。获得高质量的历史数据并不只是花钱买来 来然后装进系统那么简单。对于交易商来说，一旦他们决定交易哪种资产，第一步就是考察其数据是否可以被获得。他们要决定需要什么类型的数据，从哪里获得，如何做数据清理才能让数据适应系统。

历史数据主要有三种来源：

1.从数据提供商^Q购买数据。购买数据时，要确定数据的类型(1分钟级别的还是Tick级别的等)；还要考虑数据提供商的专业经验以及过往客户的口碑；最后，无论提供商说什么，必须自己判断数据的质量，如果自己的系统上已经有些已知的高质量数据，不妨测试一下，看买来的数据结果如何。

2.从网上下载免费的数据。这一方法并不推荐，因为你无法控制数据的质量。不可靠的数据可能导致模型计算错误，引发亏损的交易。所以相对来说，使用费用昂贵的数据反而比较划算。但即使是高质量的数据也会有问题，决不能跳过数据清理这个环节。

3.用程序记录市场行情，并将其清理和转化成想要的格式。经纪商通常会提供应用程序接口(API)来产生实时交易行情。一些商业程序化交易平台有行情写库功能，并能回放储存的数据来测试系统。这一方法要求金融工程师具有熟练的编程技巧和数据库管理能力。

数据类型

数据类型包括：

1.价格数据通常由买入价、卖出价、成交量等数据构成。对于高频交易系统来说，tick数据^Q可能是平滑数据。有些金融衍生品并没有历史成交价，相反，他们使用的是估值价格^Q。

2.经济数据一般是短期市场走势的重要因素之一，特别是在汇率市常高频交易系统对短期的市场走势敏感，所以经济数据对于系统来说比较重要。对于外汇交易来说，最重要的经济数据为非农就业指数、利率、零售业销售额、CPI、贸易顺逆差和制造业PMI等。

3.日历数据包括公布重要经济数据的日期和时间。一些高频交易系统纯粹建立在价格数据基础之上，但即使是这些系统也要对日历数据多加小心。

4.估值数据与价格数据不同。对于场内交易的衍生品来说，并没有历史价格数据。价格只存在于理论模型中，所以是估值数据。

5.基本面数据是商业上的关键数据。比如收益、销售额、库存、租金等。



下载知乎客户端

与世界分享知识、经验和见解



相关问题

[什么是高频交易系统？](#) 58 个回答[高频交易系统是不是绝大多数都是基于C++开发的？](#) 19 个回答[高频交易开发工程师如果发现有线上bug，是否必须在几分钟内解决该bug？](#) 19 个回答[什么是高频交易系统啊？](#) 1 个回答[高频交易中的性能优化是怎么做的？](#) 2 个回答

相关推荐



TradeStation 交易应用实践：量化方法构建赢家策略...

7 人读过

[阅读](#)

跨市场交易策略（典藏版）

(美)约翰J.墨菲

9 人读过

[阅读](#)

量化投资技术分析实战：解码股票与期货交易模型

22 人读过

[阅读](#)[帮助中心](#)[知乎隐私保护指引](#) [申请开通机构号](#) [联系我们](#)

数据清理过程

建立算法、历史回测和风险管理都要求高质量和处理过的数据。高频的原始数据通常伴随着市场噪音，数据过滤处理就相对比较重要。沉迷于历史回测的金融工程师往往会低估这一步骤的重要性。高频交易系统对数据很敏感，所以花些时间来处理数据至关重要。

价格数据是高频交易中最常用到的数据，一套高频交易系统依赖四个子过程来处理数据：

- 1.检查时间戳。有些时间戳的顺序并不正确，需要将它们重新排列。
- 2.排除数据坏点。寻找、代替或删除数据坏点。
- 3.过滤掉非交易日的数据。
- 4.将数据转化为不同的周期级别，如1分钟、15分钟、60分钟等，以便交易系统调用。

数据分离^Q

如果采用全部数据集来开发[交易模型](#)^Q的话，可能会出现过度拟合的问题。为了解决这一问题，我们引入数据分离过程，将全部历史数据分成开发数据、验证数据和测试数据。

1.开发数据(又称[样本内数据](#)^Q)一般包含数据最多，直接用来开发模型。数据量取决于系统的周期。一般的原则是，这段数据要产生30至50个交易样本，并且涵盖了市场行情向上、向下和振荡这几种情况。一旦确定了参数和交易策略之后，就基于开发数据来优化。

2.验证数据是用来评估和比较不同模型绩效的。如果在开发数据上表现良好的模型到了验证数据突然失败了，说明有过度拟合的问题，需要用验证数据来改进模型。

3.测试数据(又称[样本外数据](#))用来再次验证模型，并保证所有细节都正常。模型在这些数据上不再做任何改动。

建模软件编程

对开发人员来说，MATLAB、SAS和S-PLUS都是模型编程的理想工具。很多金融工程师使用Excel，但处理海量高频数据时却不从心。如果你的模型要求极高的计算性能，那你要么升级硬件，要么自主研发一套系统来支持模型。自主研发一般使用C++或C#来开发具体的软件工具，用于高速处理文件和计算。

一般来说，[模型编程](#)^Q要与实施阶段的程序完全分离，由两组人来实施。第一组人由交易员、分析师和[金融工程师](#)^Q组成，在金融建模功能工具(如MATLAB和SAS)上建立模型；第二组人由[软件工程师](#)^Q组成，他们并不对模型本身发表意见，负责根据模型要求用C++和C#语言将模型付诸实施。这一过程中，要求建模阶段要迅速，金融工程师必须尽可能快地将各种要求、设计方案和实施方案整理出来。这种分工的理由是金融工程师在量化领域可能很有优势，但未必有足够的编程能力来撰写具有高稳定性和可靠性的，用于实时交易的代码。反过来，经验丰富的程序员也未必有足够的投资知识来产生交易策略。一些交易商喜欢雇佣没有任何金融背景的纯程序员，以便他们不会随随便便在实施过程中改变交易策略。

另一种建模编程的方式叫做“[进化型建模](#)^Q”。建模团队选择系统中的一部分进行建模，然后在这个基础上扩展形成整个系统。这样建模阶段的代码就不必舍弃。相反，这些代码会“进化”成实施应用程序，用于实时交易。换句话说，实施系统是一系列模型进化后的最终产物。这种建模编程需要在一开始就做出高质量的代码。这一过程要求团队有极强的管理能力，适合那些经验丰富有高度组织性的团队。进化型建模专注于系统中风险最高、难度最大的部分，明确项目中的最大障碍有助于评估项目的可行性。这种做法通常适合那些在金融领域和软件开发领域都有不错基础的小团队。对大型团队来说，适合采用一种按部就班的项目开发流程。团队一上来先建立一个简单的模型，然后由简到繁不断增加和集成其他的系统。像MATLAB这样的建模语言能够将代码转换成C++或C#的形式，可以重复利用模型代码，实现进化型建模。

回测和优化

回测的作用一是测试交易策略是否达到了设计文档中的要求，二是基于清理过的历史数据，在考虑交易成本和滑点的情况下，对策略进行模拟仿真，并统计分析输入和输出结果。一个[交易策略](#)^Q只有在结果达到标准的情况下才被接受，否则将会重新回到研究阶段进行重做，或者直接舍弃。

举报中心

涉未成年举报 网络谣言举报 涉企虚假举报 [更多](#)

关于知乎

[下载知乎](#) [知乎招聘](#) [知乎指南](#) [知乎协议](#) [更多](#)

京 ICP 证 110745 号 · 京 ICP 备 13052560 号 - 1 ·
京公网安备 11010802020088 号 · 京网文
[2022]2674-081 号 · 药品医疗器械网络信息服务备
案（京）网药械信息备字（2022）第00334号 · 广
播电视节目制作经营许可证：（京）字第06591号 ·
服务热线：400-919-0001 · Investor Relations · ©
2023 知乎 北京智者天下科技有限公司版权所有 · 违
法和不良信息举报：010-82716601 · 举报邮箱：
jubao@zhihu.com



优化是通过改变交易策略参数达到最优结果的过程。设计人员必须发现并剔除那些在任何情况下都达不到标准的参数和交易规则。

回测和优化可以在放两个阶段内完成：建模阶段和实施阶段。建模阶段的回测和优化是为了确立或推到交易系统的设计方案。研发项目在建模阶段后决定是否继续。实施阶段的回测和优化是为了验证模型实施的有效性和质量。

发布于 2016-03-20 16:57

 [写回答](#)