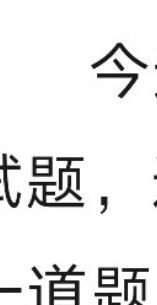


第139封信 | 股票最长的增长期



吴军



第139封信 | 股票最长的增…

09:52

4.62MB



信件朗读者：宝木

小师弟，你好！

今天再和你一同分析一道Google的面试题，这是我在Google面试时被考到的第一道题。为了便于你理解这道题，我把原题用另一种方式表述一下。

假如你研究股票投资，可能好奇一只股票最长的有效增长期是哪一个时间段，即从哪天开始到哪天结束。这样从理论上讲，如果有一个预言家告诉你这个信息，可以得到理论上最大的收益。

那么什么是有效的增长期呢？所谓有效的增长，是比大盘涨得快的那部分。我在之前很多介绍投资的信中讲过，任何一家公司最终都会消亡，因此它的股票在一定阶段后的表现会一天不如一天，直到这家公司不再存在。从那一天开始，投资它们就没有意义了。当然，由于通货膨胀，此后股票价格可能还会上涨，但并不意味着它们的价值在增加。

事实上，只有当一只股票的价格增速超过股指的时候，购买和持有它才有意义。因此，我们要扣除掉整个市场对股票价格的影响，当一只股票每天上涨的速度超过股票指数，我们就认为它的有效增长是正数，否则就是负数。为了你便于理解，我们可以看一个具体的例子。

股票ABC每日的有效增长如下：

天数	1	2	3	4	5	6	7	8	9	10	11	12	13
有效增长	1.5	-12.3	3.2	-5.5	23.2	3.2	-1.4	-12.2	34.2	5.4	-7.8	1.1	-4.9

其中每一个数字代表它今天比昨天的增长减去股票指数增长后的净值，单位是基本点，也就是万分之一。第一天它比大盘多涨了1.5个基本点，第二天又比大盘跌了12.3个基本点，以此类推。从第一天到第十三天，ABC的股票比大盘累计有效增长为27.7个基本点。注意，股票的增长其实应该做乘法计算复合增长，但是这里为了简单易懂，我就采用加法计算简单增长了，在算法上它们并没有区别。

假如有个预言家能告诉你哪天开始买入股票，哪天卖出，你就能最大地获益。从上面的数据可以看出，你应该在第5天买入，在第10天卖出，这样你的累计有效增长就比股指高出52.4个基本点。在此之前和之后，持有ABC公司的股票都不如持有指数基金。

在实际的股票交易中，思科公司和英特尔公司的股票，最佳的持有区间是从它们上市一直持有到2000年的某一天，再往后，持有它们的股票就不如持有标普500指数或者道琼斯指数了。这个最长点增长期，是一个投资人所能最大获益的上限，通常没有人能够捕捉得那么准确，但是上述的区间，对于股票研究还是有意义的，它可以作为分母来衡量投资人的表现。

接下来Google的问题是，**如何确定起始的日期和终止的日期？**

对于这道问题，有四种可行的方法，它们计算的复杂度从高到低依次下降。

方法一，做一次三重循环，其实就是中学里学的排列组合的方法。我们先假设一共有K天。

如果起始和终止日期能够确定，从起始日到终止日，做一次连加，就算出某个特定的起始日到终止日之间股票的有效涨幅，这是一重循环。当然起始日可以是第一天到第K天中的任意一天，这是第二重循环。终止日也可以有这么多选择（当然终止日要不早于起始日），这是第三重循环。于是起始和终止日本身就有 $K \times K / 2$ 种选择。每一种选择最多要做K次加法（最少做零次），平均做 $K / 2$ 次。因此这种方法的复杂度是 $O(K^3)$ ，即K的三次方。如果已经忘记了复杂度O的概念，可以翻一翻之前的内容。

这个方法无疑能够完成任务，但是做了太多的无用功。如果遇上GE这样的百年老店，数据有几万个点，几万的三次方可是几十万亿，计算量非常大。当然，这个方法稍加改进就能快很多，于是就有了下一种方法。

方法二，做两重循环。

在上述方法中，我们可以看到这样一个现象：对于某个特定的起始日期，比如从第500天开始，我们想知道终止日期设 在第1000天好呢，还是第1001天更好。这件事很简单，只要看看第1001天的有效涨幅是正的，还是负的就就可以了。如果是正的，那么显然要持有到1001天，因为多涨了一天，这时，我们要把累计到第1001天的涨幅记录下来，但是到第1000天的累计涨幅其实后面不会再用到了，就可以删掉了。

相反，如果1001天的涨幅是负的，那并不能说明股票从此就该抛掉，因为第1002天、第1003天……都还有机会大涨。在这种情况下，要保留截止到第1000天的累计增幅，然后再往后一直看，直到有一天，累计增幅超过它，才可以把累计到第1000天的数据删掉，用新的累计增幅替代它。这样一来，设定一个初始日期，扫描一遍，就能找到从这个起始日期开始，到哪一天结束，累计的增幅最大。这个过程需要K次加法和K次比较。

由于起始日期还不确定，它可以是从第一天到最后一天（第K天）中间的任何一天，因此要试K次，于是这种算法的复杂度是 $O(K^2)$ 。如果K有好几万，计算量是十几亿，比前一种方法已经减少了上万倍。但是，这个问题还有更有效的解法。

方法三，这个方法比较难理解，你只要记住这是分治算法的应用就好了。它和方法二的关系，就如同过去介绍的归并排序和冒泡排序的关系。

更具体一点讲，就是把股票涨跌的时间从中间一分为二，变成两个阶段，而两个小问题加起来也要比一个大问题简单些。然后你再把两个小问题分成四个更小的问题，直到分不下去为止。最后这个方法的复杂度可以降到 $O(N \times \log N)$ ，对于GE这样有几万天交易记录的公司，计算量为百万左右，这比十几亿又小了上千倍。然而，它依然不是最好的方法。

方法四，正反两遍扫描的方法。

这个方法也是对方法二的改进。在方法二中，我们其实能够找到股票从此走下坡路的时间，比如我们能知道在第1005天之后，股票的累计涨幅达到了顶点，再往后虽然有涨有跌，但累计下来没有增长，因此就不用考虑了，这一天就定为截止日期，或者说抛售的日期。但是，我无法知道股票在什么时候开始进入上升期的。因此在方法二中我们假设了所有的情况都是合理的。

然而，如果我们能够逆向思考这个问题，从最后一天往前回溯，用同样的办法，就能找到股票上升期的开始日期。这两个日期之间就是该购买和持有ABC公司股票的时间段。

因此，第四种算法只要扫描两遍数据，就能解决所有的问题，其复杂度是 $O(N)$ ，也就是说线性。对于GE这样有几万天交易的股票，计算量只有几万而已，比上一种方法又快了几十倍。

从方法一到方法四，我们将一个问题的计算量从几十万亿，降低到几万。平庸的方法和好的方法差距如此之大，对此不知道你对此有何感想。我在Google面试时，大约花了一刻钟时间想到了方法四，面试我的本杰明估计很满意，因为他对我的态度马上就热情了很多，而我也感叹他们的面试题出得精妙。

这道面试题唯一的缺陷是无法避免刷题，因为在找到方法四后，就很难再往深了继续提问。

最后总结一下今天的内容。

首先，我再次向你展示了**计算机科学中经常会遇到的量级的差别**。

其次，在金融上，任何一只股票都有**跑不赢大盘的时候，因为任何公司最终会消亡**。道琼斯最早的12只成分股公司和随后的20个公司，今天只剩下GE一家了。而且如果不出意外，GE的股票从此也很难再跑赢大盘了。因此，玩单只股票其实是很危险的事情。

当然，最关键的是，**逆向思维永远重要**。

关于股票投资大家可以参考我在《硅谷来信》中第145封信到第149封信的内容。

最后，如果你是计算机专业的人，希望今天的方法能够有助于你分析更多的算