浅析近十年来中国信息学竞赛知识体系的变化

周欣（2021010695，zhouxin21@mails.tsinghua.edu.cn）

1. 引言

**信息学奥林匹克竞赛（英文名：Olympiad in Informatics，之后我们用OI**作为“信息学竞赛”一词的简称**）**是一门在中学生中广泛开展的学科竞赛。在中国，信息学竞赛是官方认可的五大学科竞赛之一。2010年末，教育部开始自主招生改革[[1]](#footnote-2)，逐渐取消OI省赛一等奖保送资格。从那时起，OI发生了诸多变化，并逐渐成长为今日人们所熟悉的样貌。这其间，OI的知识体系也持续发生着变化。圈内许多人认为OI的知识体系在不断膨胀，并对这种现象予以了负面评价。对于这一看法，本文将对其前提——OI中知识体系变化的情况加以探究，并为现役选手的训练、试题的命制以及组织方的决策提供借鉴。

二、OI的基本模式

（一）赛事流程

OI赛事活动的组织者为中国计算机协会，简称CCF。每一年的OI比赛都是从十至十一月的联赛开始[[2]](#footnote-3)。联赛分省举办，包括初赛和复赛，通常一年的联赛一等人数为一至两千人。之后是省队选拔赛，入选省队的选手将获得参加国赛的资格，通常包括在每年的三至五月举行[[3]](#footnote-4)，赛题难度多与国赛持平而上下浮动。国赛是在次年七月，参赛选手约为四百人，其中五十名优胜者将入选国家集训队，获得保送大学的资格。在国赛之后还有流程冗长的持续一年的选拔国家队的赛事，直到后一年暑假前会选出四名国家队员参加国际信息学奥林匹克竞赛[[4]](#footnote-5)。

（二）比赛形式

OI比赛主要形式是上机编写代码解决特定问题，比赛时长通常为3.5至5小时，试题数量从3到4道不等，依不同比赛而定。每题满分为100分。每道题都有多个测试点，选手根据每道题的提交代码通过的测试点的数量获得相应的分数。

（三）教学方式

OI选手的知识渠道主要分为三类，在线题库与群组社区[[5]](#footnote-6)、学校内部[[6]](#footnote-7)、培训机构。并且在这三个渠道中，主要教学者基本均由现役选手或退役不久的前选手组成，职业教学人员占比很低。

（四）比赛的试题命制

OI比赛试题的命题人完全由已进入大学的前OI选手组成[[7]](#footnote-8)，但是在命题人选的确定上CCF加以监管。在过去的几年里，征题模式对题目具体内容的限制较为宽松。在2021年CCF出台了NOI考试大纲[[8]](#footnote-9)，但是相关内容还在进一步完善，仅从目前看来，大纲对于题目内容的限制仍然偏弱。

1. OI知识体系变化的情况
2. 概述

关于OI的教学内容，圈内没有形成完全统一的共识。加上教学模式的因素，实践中教学者在教学内容的选择上常常具有极大的自由度，OI的教学内容近年来一直表现为开放态势。

现行OI命题体制下命题人员的流动性很强，多数命题人为退役不到三年的前OI选手。而官方的硬性约束，以及OI圈内对于知识内容的共识所产生的隐性约束也偏弱。所以，相应的，目前运行的命题体系，在赛题内容的选取上也呈现出相当开放的态势，命题人员自由度较高。

OI中的教学者与命题者有很高的自由度，教学体系和命题体系有很高的开放性。在这样的背景下，和相对固化的高考知识体系相比，OI的知识体系更容易随着时间推移而发生变化。但是与此同时，界定OI的知识体系也变得困难了。因为OI的教学活动是高度分散而不统一的，我们很难直接通过教学活动中传递的知识来界定OI知识体系。通过正式赛事的考察内容来界定OI知识体系也是不妥的，事实上许多正式比赛考察的知识并不来自于OI知识体系，而是来自于命题人大学课程所学[[9]](#footnote-10)。一个折衷的办法是，我们将目光转向知识的接收者，通过各种手段评定选手对于知识的掌握情况，来评估OI的教学体系，进而评估OI知识体系变化的情况。

知识被选手掌握情况的变化可以分为两类，一是知识引入与普及的，二是知识的倒退。知识的引入是说，一个知识在OI的知识体系中从无到有，这自然是知识体系的变化。知识的普及是说，掌握一个知识的选手变多了。显而易见的是，从国家队选拔到联赛等不同层级的赛事，由于选拔目标、赛题难度等的差异，其教学与考察的知识体系也是不同的。如果一个知识的普及面的数量级变化到达了一定程度，说明教学体系中将该知识置为了更低的赛事层级。相应的，更低赛事层级所囊括的知识量也就增加了。知识的倒退则是说，一个知识的普及面发生了缩小的现象，说明教学体系在逐渐将该知识置为更高层级。普及与倒退都对应着知识所处层级的变化，是知识体系变化的一种表现。

想要观察OI中知识体系的变化，除了可以观察选手对知识点的掌握以外，还可以观察知识点之间的联系。教学体系的运作与知识体系的构建有其内在的规律。一个好的知识体系，其各个知识点之间的联系是紧密的，脉络是清晰的，具有较强的系统性。这样，选手才能灵活运用已有知识解题，而在面对新的问题时清楚地知道既有知识的边界在哪里。另外，对于同时充当着一线教学者的选手和前选手而言，好的知识体系也更有利于教学活动的开展。

1. 知识的引入与普及——以快速傅里叶变换为例

快速傅里叶变换，简称FFT，是目前多项式乘法问题的最优算法，曾在2014年浙江省省队选拔赛中出现。在比赛中仅有六人通过此题，而其余人由于未掌握该算法，故仅获得了朴素解法的分数。作为对比，那年浙江省总共有12人入选国家集训队。如果认为来自浙江省的集训队员与来自其他省的集训队员，对FFT的掌握情况服从同一分布，则全国范围内仅有一半的集训队员掌握该算法。

在随后的数年中，FFT的普及面快速扩大。到了2018年的清华大学信息学夏季体验营，再次出现了一道运用FFT的试题。对于该题得分情况，出题人评论，与三年前FFT仅有集训队员掌握的局势相比变化很大[[10]](#footnote-11)。而到了现在2021年，在一个名为uoj的在线评测平台上，该算法对应的经典习题[[11]](#footnote-12)，通过人数已经增长到了惊人的2500，超过了每年的联赛一等人数。而从2020年9月27日至2021年10月1日期间，uoj上该题的通过人数为227人（去重），考虑到选手的竞赛生涯通常为两年，而且只有一部分选手使用uoj，所以实际掌握人数只会更多。

FFT算法从2014年到2021年的普及，大致经历了从集训队级别，到国银级别，再到介于国赛与联赛之间级别的过程。在这一过程中，我们直观地感受到了一些知识点在OI的知识体系中变化的过程，从无到有，从难度层级较高的位置下降到较低的位置。

FFT算法当初被OI界引入的动机与具体过程现如今已难以考证。但是，包括FFT在内的诸多知识点，在近十年来确实存在着被引入与快速普及的情况。在这些热门方向上，OI知识体系的变化，主要表现为不同赛事层级的知识量均快速膨胀，知识普及面快速扩大。

（三）知识的倒退

仍然以2018年的清华大学信息学夏季体验营为例。前面提到，参赛选手对于FFT的掌握情况很好。然而命题组也了解到，存在参赛选手想到借助差分约束算法解题，但却不会编写差分约束算法代码的情况 [[12]](#footnote-13)。需要说明的是，差分约束在2016年之前的洛谷网站为用户设置的一份题单中被列入联赛级别，可见在OI界的传统认知里，差分约束并非高层级的知识点。与此同时，参加清华大学信息学体验营的基本资格是获得联赛一等奖，在实际操作过程中门槛往往会更高[[13]](#footnote-14)。

差分约束算法代码实现难度并不大，前面提到的传统认知与实际掌握情况的脱节，只能归结于新一代选手所掌握的知识体系的变化。到了2021年的全国统一的省队选拔赛中再次出现考察差分约束算法的试题时，根据各省得分的得分分布表，全国仅有寥寥数人通过该题。

（四）知识的体系化——以以近年来生成函数领域知识的演进为例

上一节提及的差分约束，被淡忘的原因并不仅仅是OI教学体系的怠惰，也与差分约束在知识体系中的位置相对孤立有关。想要掌握某个知识点，一个好的方法是积极建立其与其它知识点的联系，多练习综合运用知识的能力，将知识体系建立成牢固的网络。在这一需求的驱动下，近年来OI的知识体系除了增长速度快以外，还出现了系统性变强，知识理解程度加深的情况。本小节将以生成函数为例说明这一现象。

生成函数是将组合对象代数化的手段，现已发展为OI中组合计数领域的重要工具。

在2014年Picks’s Blog[[14]](#footnote-15)中出现了多项式求逆算法，而另外一篇博客[[15]](#footnote-16)声称这是OI中首次正式引入。在随后的一年里，通过多项式牛顿迭代方法的确立，多项式开根、对数、指数、求幂也被逐渐引入。

到了2015年集训队论文[[16]](#footnote-17)中，金策同学明确将形式幂级数作为代数对象来处理组合问题，总结了既有的形式幂级数算法的同时还引入了拉格朗日反演的技术[[17]](#footnote-18)。而吕凯风同学则确立了集合幂级数的概念，并借助子集并卷积的高效算法解决了一些问题[[18]](#footnote-19)。

此后的几年里，运用生成函数求解的组合问题逐渐增多，多类相关问题的系统性解法也在逐渐积累。2017年的集训队中毛啸同学引入了BM算法，可以从线性递归数列还原出对应的线性递推式[[19]](#footnote-20)。2017年国赛第一天的第三题需要运用多项式取模算法来优化线性递推计算。2019年的集训队论文中钟子谦同学引入了整式递推数列的概念[[20]](#footnote-21)。此外2019年的冬令营、浙江省选、国家队选拔赛中连续三次均出现了使用生成函数求解的试题，同年国赛里也出现了两道需要使用多项式插值、下降幂多项式等相关知识的试题。

2020年开始OI中的生成函数进入了一个新的发展阶段。整式递推技术得到了进一步运用，很多过去只能使用组合方法求解的递推式现在可以更系统化甚至由程序机械化地求解；很多问题得到复杂度上的改进，例如2019年国家队选拔赛中的一道试题。拉格朗日反演技术也得到了进一步发展，例如在2020年浙江省选中出现了一道需要灵活地综合运用拉格朗日反演、容斥原理、多项式牛顿迭代等的试题。

此外，这一阶段仍然在引入新技术，例如2020年集训队论文中陈宇同学介绍了转置原理，并展示了如何运用转置原理在多项式多点求值问题中避免多项式除法[[21]](#footnote-22)。例如这一时期组合符号化方法的引入，并对2018年浙江省选中出现的一道试题给出了比官方题解更易懂的解释。又比如李白天同学在他博客中[[22]](#footnote-23)介绍的线性递推的新方法，规避了多项式除法的使用，并且运行效率和实现难度都得到了提升。这一阶段的集大成者是李白天同学，感谢他的工作。

可以看到，生成函数在过去几年的发展，除了知识点的简单增多以外，知识点之间的联系也更加紧密，更具系统性了。选手们对于旧有知识点的理解与运用更加深刻，看待问题的观点与视角更高更清晰，知识创新的层面上也取得了一些令人欣喜的成就，甚至出现了多项式除法应用面大幅倒退这种令人诧异的情况。

同时我们也可以重现审视一下第二小节中提到的FFT，作为目前多项式乘法问题的最优解法，其在整个生成函数理论中处于基石的位置。FFT的普及，实际上是与整个生成函数计算理论的快速发展相适应的。

事实上除了生成函数领域，OI的其它知识领域，例如数据结构、字符串、组合优化等，也或多或少出现了知识体系快速膨胀的同时理解程度加深、系统性变强的情况。这种现象并不是生成函数领域的特例。

四、结语与展望

OI由于其运行模式的固有特点，其知识体系近年来在某些热门方向上确实呈现了膨胀的现象，但同时在一些冷门方向上也存在着知识倒退的现象。从另一个方面说，当我们站在知识的系统性的角度来看待问题时，可以发现OI的知识体系确实存在着系统性变强、理解更深入、运用更灵活的现象。

此外，在本研究的探索过程中，还有许多新的话题得到发掘。比如，在自治为主而缺乏外力监管的情况下，OI知识体系的变化有着怎样的内在机制和规律。OI的教学与命题体系，OI的网络平台社群及其承载的文化，具体是怎样塑造着OI的知识体系的呢？ 又比如，目前这种知识体系持续变化状态的存在，从不同视角看分别有哪些利与弊呢？值得思考的问题还有很多。

1. 《教育部：关于调整部分高考加分项目和进一步加强管理工作的通知》， 2010年12月13日，https://gaokao.chsi.com.cn/gkxx/ss/201012/20101213/149734454.html [↑](#footnote-ref-2)
2. 《CCF关于NOIP 2021的报名通知》，2021年10月23日，https://noi.cn/xw/2021-10-23/746061.shtml [↑](#footnote-ref-3)
3. 《CCF关于NOI2021省内选拔的若干规定》，2020年12月24日，https://www.noi.cn/xw/2020-12-24/719916.shtml [↑](#footnote-ref-4)
4. 《全国青少年信息学奥林匹克竞赛系列活动简介》，2020年9月29日，https://www.noi.cn/gynoi/jj/2020-09-29/708625.shtml [↑](#footnote-ref-5)
5. 李白天：《网络社群的发展对信息学竞赛的影响》，2021年 [↑](#footnote-ref-6)
6. 徐西玲：《整型溢出：信息学竞赛的发展，繁荣与衰退》，2020年，第21-22页 [↑](#footnote-ref-7)
7. 徐西玲：《整型溢出：信息学竞赛的发展，繁荣与衰退》，2020年，第32-33页 [↑](#footnote-ref-8)
8. 《NOI大纲正式发布》，2021年4月2日， <https://www.noi.cn/xw/2021-04-02/724387.shtml> [↑](#footnote-ref-9)
9. 王天懿：《论偏题的危害》，https://github.com/OI-wiki/libs/blob/master/topic/7-%E7%8E%8B%E5%A4%A9%E6%87%BF-%E8%AE%BA%E5%81%8F%E9%A2%98%E7%9A%84%E5%8D%B1%E5%AE%B3.ppt ，2015年NOI冬令营营员交流 [↑](#footnote-ref-10)
10. 如何评价 THUSC 2018？ -知乎，<https://www.zhihu.com/question/279665287/answer/409243750> [↑](#footnote-ref-11)
11. <https://uoj.ac/problem/34> [↑](#footnote-ref-12)
12. 如何评价 THUSC 2018？ -知乎，<https://www.zhihu.com/question/279665287/answer/409243750> [↑](#footnote-ref-13)
13. 徐西玲：《整型溢出：信息学竞赛的发展，繁荣与衰退》，2020年，第50页 [↑](#footnote-ref-14)
14. <http://picks.logdown.com/posts/189620-inverse-element-of-polynomial> [↑](#footnote-ref-15)
15. <https://blog.csdn.net/VFleaKing/article/details/90521383?spm=1001.2014.3001.5501> [↑](#footnote-ref-16)
16. CCF：《关于IOI2021国家队选拔流程的说明》，2020年10月10日，https://www.noi.cn/xw/2020-10-10/715672.shtml [↑](#footnote-ref-17)
17. 金策：《生成函数的运算与组合计数问题》，2015年国家集训队论文 [↑](#footnote-ref-18)
18. 吕凯风：《集合幂级数的性质与应用及其快速算法》，2015年国家集训队论文 [↑](#footnote-ref-19)
19. 毛啸：《关于数列递归式的一些研究》，2017年国家集训队论文 [↑](#footnote-ref-20)
20. 钟子谦：《两类递推数列的性质和应用》，2019年国家集训队论文 [↑](#footnote-ref-21)
21. 陈宇：《转置原理的简单介绍》，2020年国家集训队论文 [↑](#footnote-ref-22)
22. 李白天：《一种新的线性递推计算方法》，2020年10月21日， <https://blog.csdn.net/EI_Captain/article/details/109196620?spm=1001.2014.3001.5501> [↑](#footnote-ref-23)