**1. 优先最长学科代码动态专家匹配方法**

**国家自然科学基金申请代码**

国家自然科技委员会依据科学部的不同，将自然科学基金划分了 8个部分，分别使用 A、B、C、D、E、F、G 和 H 来表示所属科学部，而每一个大的科学部又逐步划分更加细致的学科类别，层层划分、使用数字进行编码。

如：F010703：水下通信网络

优先最长学科代码动态专家匹配思想是基于每一个待评审的科技计划项目都有唯一对应的学科代码。基于该代码在专家库中优先寻找与之相同的学科代码，若没有匹配到完全相同的，则将该学科代码回退两位，继续在专家信息库中匹配与该代码一致的专家，如此循环，直到查找到足够数量的评审专家。

流程图解释：

1. 首先检测的是总的数目和学科代码长度是否大于 3，检测总数是基于项目总评审专家数量规约。而学科代码长度阈值为 3 是基于国家自然科学基金委员会学科代码定制规则，当学科代码长度少于 3 的时候，其所表示的范围已经非常广泛了，匹配所得结果集已经很大，该结果集对科技项目管理人员而言并无太大参考意义。

2. 其次，对于同一批次的项目评审过程中，必须设定每个专家评审项目的上限，因为人的精

力是有限的。

3. 规避与公平：

规避原则在于：首先自己申请的项目，评审专家不可以是自己。

公平性在于：必须设定一定比例的省内外专家人数，以避免同地区互熟，对评审结果造成影响。

产出与结果例子：例如专家评审过的项目 P，申请人所填写的对应的项目学科代码为 F020403、F020201，而通过使用本节所提出方法所得评审专家：甲、乙、丙、丁、戊各自所填写的研究学科代码均有 F020403 以及 F020201。

一些可能的问题

但是还会存在另外一个问题。例如专家 A 填写的可评审学科代码表示的是计算机应用技术，其实际研究点在于深度学习、人工智能、模式识别等细分研究点。而申请的科技计划项目所填写的学科代码也表示计算机应用技术，但是具体研究医学数据分析与处理，该项目是着力使用计算机技术来处理医学领域研究的课题，会涉及很多的医学知识背景。然而，依据本文使用的优先最长学科代码专家匹配机制是可以将二者做评审匹配的。依据专家同行评议原则，该匹配是合理的，但是，从挑选评审专家的角度来说，对于科技计划项目评审专家的挑选不仅希望研究领域要确保准确，更加希望领域下细分的研究点也是一致或者相近的。

**2. 关联规则挖掘**

**支持度**

同样“A+B”出现了 3 次，那么这 5 组中“A+B”的支持度就是 3/5=0.6

**置信度**

置信度（E，A）=2/3=0.67，代表组里有E时，有多大的概率会有A？

置信度是个条件概念，就是说在 A 发生的情况下，B 发生的概率是多少。

**提升度**

提升度 (A→B)= 置信度 (A→B)/ 支持度 (B)

这个公式是用来衡量 A 出现的情况下，是否会对 B 出现的概率有所提升。

所以提升度有三种可能：

提升度 (A→B)>1：代表有提升；

提升度 (A→B)=1：代表有没有提升，也没有下降；

提升度 (A→B)<1：代表有下降。

**3. 协同过滤**

其中公式中：r0表示待评审项目与匹配专家的相似度，sim[pi，hi]表示待评审项目 与历史评审项目h 的相似度，W 表示权重，prof 表示待评审项目pi匹配的专家，PROF ， 表示相似项目hi所对应的历史评审专家列表。

最后，可以根据r0排序出r0从大到小的专家