李熙：对通用智能模型AIXI的理论基础的研究（2015）

Hutter[17]认为，AIXI提供了“智能”本身的第一个自上而下的、有效的、广泛的、动态的、无偏的、客观的、“无参数”的理论基础。本文详细研究通用智能模型AIXI 的理论基础，除了大家普遍关注的“逼近”问题外，所谓的“无参数”的理论模型事实上仍留下了很多问题。

本文首先从哲学角度分析AIXI 的背景假设及其后续发展所面临的问题，然后从期望误差界的角度分析为什么用贝叶斯混合解决通用归纳问题而不是极小描述长度原则；通用归纳是通用智能的核心，通用先验又是通用归纳的核心，通过分析贝叶斯混合、贝叶斯博弈的理论基础将会把问题的关键聚焦在通用先验的选择上。本文从第二章开始对通用先验的合理性问题展开研究。首先从指标系统的角度提出一种备选方案并比较其与Solomonoff 先验的优略，然后从奥卡姆剃刀、最大熵原则、最优编码、期望误差界/平均冗余/信道容量等角度对通用先验的理论基础进行分析，提出可供选择的先验。最后在一种弱的意义上提出对Solomonoff 先验的可能的辩护；一个好的先验的标准之一是能确认“所有乌鸦都是黑的”这种全称命题，卡尔纳普的归纳逻辑被批评的原因之一就是不能确认全称命题，但要用Solomonoff 算法概率确认全称命题就必须对语言进行扩充。第三章通过嫁接Solomonoff 通用归纳与卡尔纳普的归纳逻辑给出了一种微弱的扩张。卡尔纳普公式是一种平滑方法，类似的比如图灵估计、线性插值等平滑方法也一样不具有通用性。本文将Solomonoff “先验”引入归纳逻辑，可以用这种修改后的归纳逻辑针对某种具体模式做出预测，还可以用它处理全称命题的确认问题，甚至可以通过一种随机抽样的方法确认全称命题；Hutter 的通用智能模型AIXI 可以看作通用归纳加动态规划的组合，本文第四章通过定义两种“可观察行为的贝叶斯博弈”发现，只要采用“通用先验”进行Harsanyi转换，那么，AIXI及其相关的策略对应这两种游戏的Ex Post 均衡、贝叶斯-纳什均衡、完美贝叶斯均衡。但AIXI 也有很多局限性。第4.2节借助纽康姆问题讨论了AIXI 作为“可观察行为的贝叶斯博弈”的最优主体所面临的“时序”不确定或说“游戏设定”本身给出的先验信息怎么合理利用的高阶不确定性问题。有了通用AI的理论，对于通用AI与具体AI的关系，第4.3节讨论了从通用AI 视角下如何看待具体AI 的问题。具体AI 问题是在资源不足的情况下故意“选择性遗忘”、“合理”划分信息集、从而生成易处理的不完美信息博弈的结果。可惜无法借助通用AI 的算法概率自动处理真实的状态转移，无法独立于环境事先自动判断哪些信息是“无用的”、“可忽略的”。“智能”的概念甚至可能远比AIXI 的框架复杂，这个框架仍存在一些值得商榷的参数问题，比如，贴现函数的时间协调性问题，效用函数不确定时的决策问题等等。第4.4.3节还揭示，用来预测的通用贝叶斯混合和衡量智能的Ex Interim 期望累积效用都依赖于通用先验，二者如果不是基于同样的先验则会面临“价廉未必物美”的问题，“简单性”未必带来“有用性”；效用还是简单性，这是一个困难的选择。第五章首先讨论了在有一个客观的外在效用的情况下，具有“乐观主义”倾向的主体可能的策略，定义了外

在效用引导的通用先验和“实用主义”的理性主体。如果没有客观的外在效用，那么借助某个合适的“内在效用”引导学习的过程就是合理的。对于如何定义“内在效用”，本文根据莱布尼茨的“完满性”哲学提出了一种直观上合理的方案，最后根据莱布尼茨“完满性”原则给出一种将“简单性”、“似真性”与“内在有用性”统一起来的通用先验。最后一章是对本文内容的总结，以及对莱布尼茨的整体哲学规划框架的历史回顾和对以AIXI 为技术框架的通用智能模型发展道路的展望。