研究揭示短时间观测数据的未来预测理论和方法

作者：writer 来源：中国科学院

本文原地址：https://www.iikx.com/news/progress/2563.html

本文仅供学习交流之用，版权归原作者所有，请勿用于商业用途！

研究揭示短时间观测数据的未来预测理论和方法。10月8日，国际学术期刊《美国国家科学院院刊》(PNAS)在线发表了中国科学院生物化学与细胞生物学研究所陈洛南研究组与苏州大学副教授马欢飞、东京大学教授Kazuyuki Aihara、复旦大学教授林伟合作的题为Randomly Distributed Embedding Making Short-term High-dimensional Data Predictable(《高维短序列数据预测的随机嵌入分布方法》)的最新研究成果。该成果提出了基于非线性动力学的全新随机嵌入理论和方法(随机嵌入分布方法RDE: Randomly Distributed Embedding)，通过低维嵌入映射获得目标变量预测值的分布，最终使得高维短序列时间序列数据的预测成为可能。即建立了由短时间观测的高维数据，预测目标变量动态行为的全新理论和方法。

在时间序列分析中，一般认为在获得低维系统的大量时间样本(时间序列数据)后，系统的重构或者预测是可行的，而短的时间样本数据一般不可预测。但是在大数据时代，人们往往获得大量的变量和有限的时间样本(如影像数据或组学数据)，一方面高维变量使得系统的拟合所需要的参数快速增长带来维度灾难，另一方面较短的时间样本往往不能获得完整的系统动力学行为或统计规律，这就对数据分析方法提出了新的挑战。

基于这个问题，研究人员建立全新随机嵌入理论 (RDE)和短时间数据预测方法，使用低维嵌入映射来构造弱预测器，在大量弱预测器的基础上构造强预测器，从而避免了维度灾难，并由高维系统中不同变量间的交互作用构建目标变量的动态信息，弥补了短时间样本的信息不足。研究人员从理论上给出了该框架的可行性分析，并通过基因表达数据、空气污染、疾病数据与气象数据等实际数据的预测进一步验证了该方法的可行性和优越性。特别是，随机嵌入分布方法转换高维数据为目标变量(低维)的动态信息，实现短序列数据的预测，同时RDE可看作是由高维小样本构建目标变量的大样本数据方法(见图1)。

即使学习数据只是吸引子的一部分样本(黑色)，但RDE可预测那些没有学习过的动态行为(红色)，见图2。

该工作对于大数据的分析，特别是高维短序列的时间序列数据分析提供了全新的概念和理论，不仅可用于时间序列的预测，也可应用于人工智能及脑科学中的大样本数据构建和全新学习建立等。

该研究得到中科院B类先导专项、国家重点研发计划和国家自然科学基金的经费支持。

文章链接

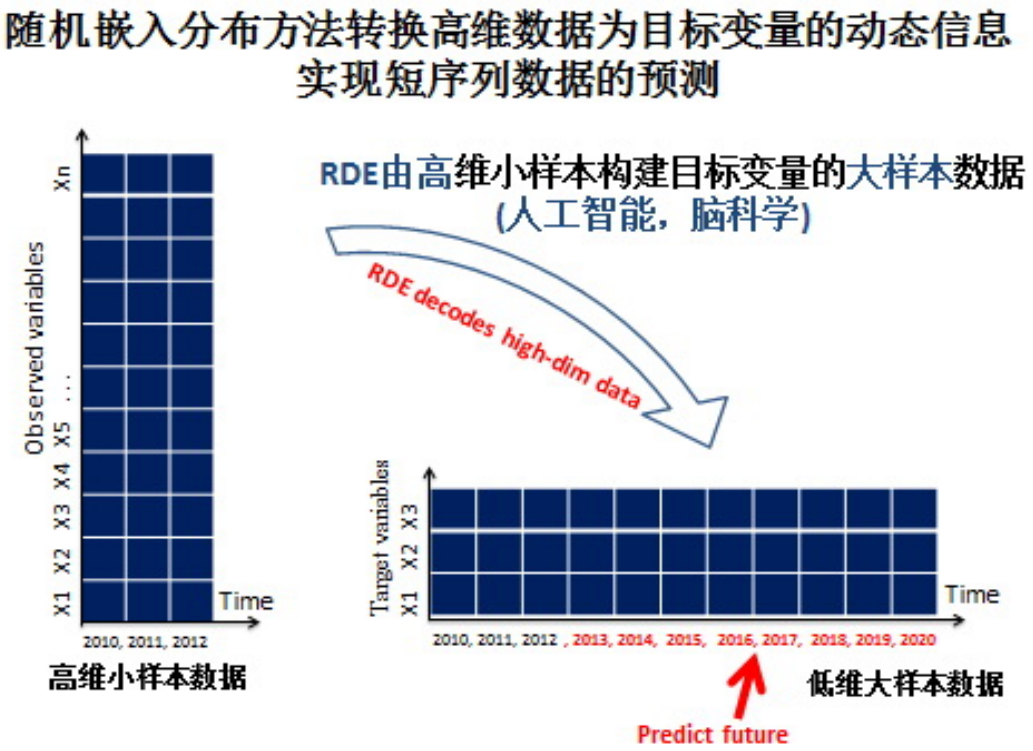


图1 随机嵌入分布方法转换高维数据为目标变量(低维)的动态信息，实现短序列数据的预测，同时RDE可看作是由高维小样本构建目标变量的大样本数据方法

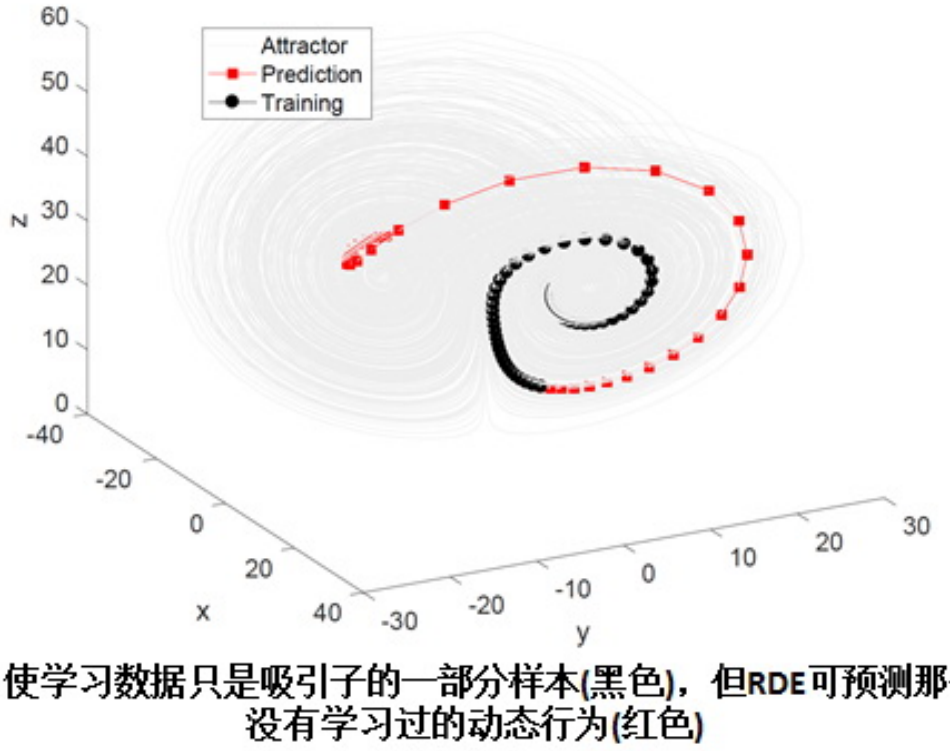


图2 即使学习数据只是吸引子的一部分样本(黑色)，但RDE可预测那些没有学习过的动态行为(

红色)

更多 科学进展 请访问 https://www.iikx.com/news/progress/

本文版权归原作者所有，请勿用于商业用途，爱科学iikx.com转发