

基于深度学习算法的中国干散货港口的上下游关联信息挖掘

刘旭锋、韦春竹

中山大学；南方海洋科学与工程广东省实验室（珠海）

摘要：基于深度学习算法的中国干散货港口的上下游关联信息挖掘 海运承担了全球约 80% 的运输任务，占全球运量比例超 71%。面对全球贸易不断重塑的背景，积极整合全球海运的时序大数据，科学评估我国港口在全球航运网络中的功能和地位，对于积极引导我国港口规划与布局、航线动态调整与变化具有重要意义。已有的港口时空关联关系研究大多基于航线较为固定的集装箱班轮运输网络，对于干散货运输的港口时空关联关系缺乏系统研究，其主要原因是干散货港口的船舶船期不定、航线多变、周期性往返频率较高，如何从多源异构的干散货船只轨迹中重构贸易网络并挖掘港口的时空关联关系仍是一个挑战。因此，本研究拟首先结合大宗货物贸易流数据、AIS 数据等，利用 2019-2020 年高时空分辨率的航运数据，重构原油、煤炭、铁矿石三类货物航运贸易网络；其次，采用 Port2vec、Seq2Seq、graphGAN 三种深度学习算法，对干散货船只的时空轨迹序列进行高维聚合，挖掘中国干散货港口的上下游信息。最后，对比深度学习算法和传统静态网络中港口上下游关系的相似与差异，总结中国干散货港口与全球其他干散货港口的潜在时空关联关系。研究表明： 1、全球间接参与干散货港口的比例不容忽视。2020 年全球直接参与原油、煤炭和铁矿石三类航运的港口总数为 1,552，分别有 655、856、639。直接参与的专业性和综合性港口分别占比 40.7%，26.2%。其中，中国三类货物直接参与的港口总数为 152，综合性港口占比 19.7%，原油港口以综合性港口为主，煤炭和铁矿石港口以专业性港口居多。此外，全球间接参与(存在船舶停靠，而无货物装卸)的港口总数为 2,772，专业性和综合性港口分别占比 3.4%，82.2%，这表明较多的港口虽未直接参与到航运之中，但其提供船舶中转、船舶暂时停留需求等功能，这些港口的重要性不可忽视。 2、干散货港口的时空序列模拟预测受到时间窗口的影响较大。研究使用 Seq2Seq 模型，以 2019 年的停港序列作为训练集，预测 2020 年的首个停靠港和长时停港序列。结果表明 Seq2Seq 模型对首个停靠港的预测效果较差，仅为 6%-10%，而预测的长时停港序列与原始序列的语义相似度在 60%以上。这表明 Seq2Seq 模型在某些情况下可能更适合预测长期趋势而不是单个事件。在港口停港序列中航线和停靠港之间的复杂关系可能导致模型在预测首个停靠港时的性能较差，而长时停港序列可能具有更强的序列模式和统计规律，使得 Seq2Seq 在预测长时停港序列时表现更好。 3、不同类型干散货港口的贸易网络模拟预测差异较大。本研究以使用 Port2vec 和 graphGAN 模型对于干散货港口的贸易网络进行模拟预测，其中 Port2vec 对原油、煤炭和铁矿石网络的预测结果分别为 48%、37%、50%，graphGAN 对三类货物的边预测准确率均可达 60%。综上所述，干散货船只的时空序列和网络结构受到货物类型和运输时长的影响具有较大的不确定性，未来研究还需要结合更多高时空粒度的船只运输信息和港航联盟信息，来提高对于干散货港口时空关联关系的模拟分析。

关键词：航运大数据、时空网络、AIS、深度学习