

DOI:10.16356/j.1005-2615.2024.04.017

基于复杂网络的空中交通流量短期预测

王飞¹, 魏林琳²

(1. 中国民航大学空中交通管理学院, 天津 300300; 2. 黑龙江省机场管理集团有限公司哈尔滨太平国际机场飞行区管理部, 哈尔滨 150079)

摘要: 为合理预测空中交通流量, 结合复杂网络链路预测进行研究。首先, 将时间序列转化为可视图得到拓扑特征量, 然后结合基于局部信息、路径和随机游走的算法, 比较在三亚不同扇区内的预测精度, 发现 RWR0.85 算法预测精度最高。由于链路预测只能预测可能存在的连边, 不能预测节点, 因此引入 D-S 证据理论预测流量值, 预测精度最高可达 99.85%。结果表明, 复杂网络链路预测结合 D-S 证据理论进行空中交通流量的预测是可行有效的, 为进一步深入研究奠定了基础。

关键词: 复杂网络; 空中交通流量; 链路预测; 时间序列; D-S 证据理论

中图分类号: V355.1 **文献标志码:** A **文章编号:** 1005-2615(2024)04-0741-09

Short-Term Prediction of Air Traffic Flow Based on Complex Network

WANG Fei¹, WEI Linlin²

(1. College of Air Traffic Management, Civil Aviation University of China, Tianjin 300300, China; 2. Harbin Taiping International Airport Flight Area Management Department, Heilongjiang Airport Management Group Co. Ltd., Harbin 150079, China)

Abstract: To predict air traffic flow reasonably, research is conducted in conjunction with complex network link prediction. First, the time series is converted into a visual graph to obtain the topological feature quantity, and then combined with the algorithms based on local information, path, and random walk, the prediction accuracy in different sectors of Sanya is compared. It is found that the RWR0.85 algorithm has the highest prediction accuracy. Due to the fact that link prediction can only predict possible edges and cannot predict nodes, the D-S evidence theory is introduced to predict traffic values, with a maximum prediction accuracy of 99.85%. The results indicate that the combination of complex network link prediction and D-S evidence theory for predicting air traffic flow is feasible and effective, laying a foundation for further in-depth research.

Key words: complex network; air traffic flow; link prediction; time series; D-S evidence theory

流量预测一直是空中交通流量管理领域研究的热点问题。空中交通流量预测一般分为中长期和短期预测。其中, 短期预测是对空域运行单元未来 6 个小时内流量进行预测, 有助于流量管理人员及时掌握未来空域运行状况, 减少空域拥堵和航班

延误, 为预先制定相应疏导策略和管制指令提供支持依据^[1], 对实际运行更具指导意义。

空中交通系统是复杂非线性系统, 近年来兴起的复杂网络是分析复杂系统的有效工具。复杂网络方法将时间序列映射成网络, 将时间序列预测问

基金项目: 天津市应用基础多元投入基金重点项目(21JCZDJC00840); 中央高校基本科研业务费专项资金项目(3122019129)。

收稿日期: 2023-03-28; **修订日期:** 2023-08-17

通信作者: 王飞, 男, 副教授, E-mail: feiwang@cauc.edu.cn。

引用格式: 王飞, 魏林琳. 基于复杂网络的空中交通流量短期预测[J]. 南京航空航天大学学报, 2024, 56(4): 741-749.
WANG Fei, WEI Linlin. Short-term prediction of air traffic flow based on complex network[J]. Journal of Nanjing University of Aeronautics & Astronautics, 2024, 56(4): 741-749.

题转化为链路预测问题。链路预测是指通过已知的网络拓扑结构以及网络节点属性等信息,预测网络中尚未产生连边的两个节点之间产生链接的可能性^[2-3]。链路预测通过掌握时间序列数据的发展过程,推理其本身的趋势和未来的走向,可以帮助人们提前规划、预防、观测,具有相当高的实际应用价值,且能够更好地反映时间序列复杂网络的动力学演变与传播过程,能够更好地了解网络中各节点之间的相互作用,直观地揭示复杂系统的演变与发展机理。

以往的复杂网络链路预测更多针对于实体网络或物理网络^[4-6],面向时间序列的链路预测研究还比较少。目前,在空中交通流量预测领域,王超等^[7]提出了基于改进加权一阶局域法的混沌交通流量时间序列预测模型,不仅能够有效预测交通流量且提高了预测精度,但还有待深入研究统计间隔对预测精度影响的内在机理。宋维凯等^[8]提出了基于动态时间规整和长短期记忆的空中交通流量短期预测模型,提升了流量预测结果的准确性,但模型仅利用时间相关性进行流量预测未考虑空间相关性。利用复杂网络链路预测方法的研究主要用于航路网络的生成及优化^[9-10],仅判断是否存在连边。应用复杂网络链路预测进行空中交通流量预测,不仅要判断是否可能存在连边,还要预测下一时间节点的流量值,相关研究尚未见诸于报道。

本文以扇区流量时间序列为研究对象,首先介绍复杂网络链路预测的相关概念和评价方法,并引入D-S证据理论,在复杂网络链路预测方法预测连边的基础上进行了节点的预测,最后进行实例分析。

1 复杂网络链路预测

链路预测包含3类:基于相似性的链路预测、基于最大似然估计的链路预测和运用概率模型进行链路预测^[6]。文献[11]认为网络连接反映内在层次结构,因此提出最大似然估计法研究明显层次网络的链路预测,拥有十分可观的准确率^[12]。但由于每一次预测需要生成大量样本网络,其计算具有高度复杂性,仅适用于较小的网络^[6]。概率模型方法对网络和节点的信息处理更细节,从而提高预测精确度,但其计算复杂性和不通用的参数限制了它的使用范围^[6]。而基于相似性的方法简单直接,能够很好地抓住目标网络的结构特征。因此本文采用基于相似性的链路预测方法。

1.1 基于相似性的链路预测

根据所用信息的不同,基于相似性的链路预测算法分为3类:基于局部信息的相似性算法、基于路径的相似性算法和基于随机游走的相似性算法。

(1)基于局部信息的相似性算法。需要的信息量最少,复杂性较低,但在集聚系数较高的网络中表现很好,然而对于集聚系数较低的网络则预测精度就较差。最简单的指标是共同邻居(Common neighbors, CN)^[12-13]。在CN的基础上,还有考虑节点度影响的Salton指标^[6]、分析两端节点度的Jaccard指标^[14]、适宜研究生态数据的Sorenson指标^[15]、大度节点有利指标(Hub promoted index, HPI)^[16]、大度节点不利指标(Hub depressed index, HDI)、LHN-I指标^[17]。Adamic等^[18]为每个共同邻居节点赋予权重提出AA指标,Zhou等^[19]衡量节点间的相似性提出资源分配(Resource allocation, RA)指标,以及只考虑两个端点的度的优先连接的(Preferential attachment, PA)指标^[12-13]。

(2)基于路径的相似性算法。从路径的信息角度定义了相似度,其精确度主要依据能否准确抓住目标网络的结构特征^[14]。常用的指标是局部路径指标(Local path, LP)^[20]、Katz指标^[14]和LHN-II^[17]指标。

(3)基于随机游走的相似性算法。常用的指标是平均通勤时间(Average commute time, ACT)^[6]、有重启的随机游走(Random walk with restart, RWR)^[6]、局部随机游走指标(Local random walk, LRW)^[21]、叠加的局部随机游走指标(Superposed random walk, SRW)^[22]。

1.2 链路预测精度指标

衡量链路预测算法精确度的指标有:受试者工作特征曲线下区域的面积(Area under the receiver operating characteristic curve, AUC)和Precision^[6,22]。其中,AUC从全局角度对算法精度进行度量;Precision仅考虑对于排在前L位的边的预测是否正确^[23]。

2 流量预测

2.1 D-S证据理论

由于链路预测的方法只能预测节点之间是否存在连边,而不能预测节点,不能直接预测出流量值,且仅考虑了最大相似度节点的影响,易造成偏差从而影响预测结果精确度。D-S证据理论不仅能处理随机性和模糊性所导致的不确定性,而且可以依靠证据的积累不断地缩小假设集,这些优点使证据理论的应用越来越广^[24]。因此,在链路预

测的基础上,通过结合 D-S 证据理论的方法对产生影响的相关节点进行权重划分,找出多个影响较大的相关节点进行流量预测,即预测节点。

假设辨识框架 Θ 是一个由问题的所有假设组成穷举集合,所有假设是相互排斥的,设 Θ 中包含 N 个元素, Θ 可以表示为 $\Theta = \{\theta_1, \theta_2, \dots, \theta_N\}$, Θ 的子集 A 称为命题, Θ 的幂集 $P(\Theta)$ 可以表示为

$$P(\Theta) = \{ \emptyset, \{\theta_1\}, \{\theta_2\}, \dots, \{\theta_N\}, \{\theta_1, \theta_2\}, \{\theta_1, \theta_3\}, \dots, \{\theta_1, \theta_2, \dots, \theta_i\}, \dots, \Theta \} \quad (1)$$

式中 $P(\Theta)$ 中元素的个数为 2^N 。

在辨识框架中,函数 m 被定义为一个映射: $m: P(\Theta) \rightarrow [0, 1]$, 且满足以下条件

$$m(\emptyset) = 0, \sum_{A \subseteq \Theta} m(A) = 1 \quad (2)$$

函数 m 为辨识框架 Θ 上的基本概率指派函数 (Basic probability assignment, BPA), $m(A)$ 表示命题 A 的基本概率指派, $m(A)$ 的值表示证据对命题 A 的支持程度。

给定几个同一识别框架上基于不同证据的信任函数,若证据不是完全冲突的,则可以利用 D-S 合成法则计算出一个新的信任函数。设 B_{e1}, B_{e2} 是同一个识别框架上的两个信任函数, m_1 和 m_2 分别是其对应的基本可信度分配,对应焦元分别为 A_i 和 B_j , 设 $K = \sum_{A_i \cap B_j = A} m_1(A_i) m_2(B_j) < 1$, 则 D-S 合成法则为

$$m(A) = \frac{\sum_{A_i \cap B_j = A} m_1(A_i) m_2(B_j)}{1 - K} \quad A \neq \emptyset \quad (3)$$

2.2 流量预测

设时间序列为 $(t_1, y_1), (t_2, y_2), \dots, (t_i, y_i), \dots, (t_N, y_N), (t_{N+1}, y_{N+1}), (t_i, y_i)$ 为时间序列中任意某个节点, (t_N, y_N) 为最后已知节点, (t_{N+1}, y_{N+1}) 为预测节点, 其中 y_{N+1} 是时间序列在时间点 t_{N+1} 的观测值。基于上文不同算法公式可得最后节点 (t_N, y_N) 与其先前所有节点之间的相似度为 $S_N = \{S_{1N}, S_{2N}, \dots, S_{(N-1)N}\}$, 设 S_{MN} 表示 S_N 中的最大值, 其对应的节点 (t_M, y_M) 即为已知最后一个节点 (t_N, y_N) 相似度最大的节点^[25]。

步骤 1 将原始时间序列通过可视图算法转换成复杂网络, 预测值 y_{n+1} 与已知网络中最后一个节点相连接。

步骤 2 在不同算法中找出预测精度最高的链路预测方法, 根据公式计算节点 (t_N, y_N) 与之前 $(N-1)$ 个节点的相似度, 记为 $S_N = \{S_{1N}, S_{2N}, \dots, S_{(N-1)N}\}$ 。为提高预测精确度, 利用链

路预测方法计算两次相似度, 记为 S_{N1}, S_{N2} 。

步骤 3 找出前 ω 个相似度最大的相关节点, 并根据式 (4) 依次对相似节点进行预测, 可得到 ω 个预测值。 ω 的取值参考经验规则^[26]为

- (1) 若时间序列呈现为上升趋势, 则 $\omega = 2$;
- (2) 若时间序列呈现周期性的状态, 则 $\omega = 4$;
- (3) 若时间序列呈现为下降趋势, 则 $\omega = 2$ 。

$$y_{N+1} = \frac{y_N - y_M}{t_N - t_M} (t_{N+1} - t_M) + y_M \quad (4)$$

步骤 4 取两次相似度计算出 ω 个最大关键节点的重合节点, 依据 D-S 合成法则合成。

步骤 5 根据时间节点的证据差异性和时间距离计算出前 ω 个相关节点的权重系数。

当时间序列无周期性时, 通常节点 (t_M, y_M) 离最后已知节点 (t_N, y_N) 越远, 说明节点 (t_M, y_M) 承载了更多历史信息也就越重要。相反如果节点 (t_M, y_M) 非常接近最后已知节点 (t_N, y_N) , 则 (t_M, y_M) 的重要性减弱, 因为其包含的过去信息较少, 对下一时刻的预测结果 (t_{N+1}, y_{N+1}) 产生的影响和 (t_N, y_N) 具有十分接近的效果。如图 1 所示, 随着节点 (t_M, y_M) 从 t_{M3} 移动到 t_{M1} , 到最后已知节点 (t_N, y_N) 的距离 d_{MN} 逐渐减小, 节点 (t_M, y_M) 携带的历史信息减小, 重要性随之减弱。

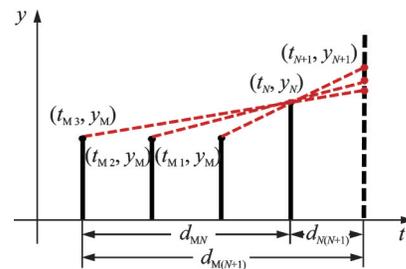


图 1 节点重要性影响

Fig.1 Impact of node importance

(1) 根据前 ω 个节点的基本概率分派, 使用式 (5) 计算其与最后一个节点的距离。

$$d_j(m_{t_M}, m_{t_N}) = \sqrt{\frac{1}{2} (m_{t_M} - m_{t_N})^T * \underline{D} * (m_{t_M} - m_{t_N})} \quad (5)$$

证据体之间的相似性计算为

$$\text{Sim}(m_{t_M}, m_{t_N}) = 1 - d_j(m_{t_M}, m_{t_N}) \quad (6)$$

(2) 两个节点间的紧密程度通过节点距离反映, 节点距离越近证明关系越紧密。这一理论应用于可视图, 越是相近的时间, 就代表着节点所含的信息越是相似, 对预测的效果也就没有太大的差别。基于此, 在权重系数的设定和预测结果的改进时, 时间距离影响是必须要考虑的。根据文献^[27], 可见性图中的节点距离定义为

$$d_{M \rightarrow N} = |t_M - t_N| \quad (7)$$

式中： t_M, t_N 分别代表对应时间节点 $(t_M, y_M), (t_N, y_N)$ 的具体时刻。

将证据相似性与时间距离这两个影响因子综合起来,得出有关的时间节点对预测结果的权重计算方法为

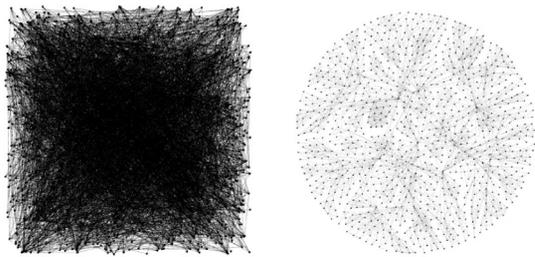
$$\omega_i = d_{M \rightarrow N} \times \text{Sim}(m_{t_M}, m_{t_N}) \quad (8)$$

步骤6 为不同的初步预测值分配权重,用来进行最终的预测,表达式为

$$y_{N+1} = \frac{\sum_{i=1}^k \omega_i}{\sum_{i=1}^k \omega_i} \times \left(\frac{y_N - y_M}{t_N - t_M} (t_{N+1} - t_M) + y_M \right) \quad (9)$$

3 实验结果与分析

为评估和检验算法,共用了3个扇区3个时间尺度的空中交通流量数据,分别为:三亚1、2、4扇区的15、30和60 min流量数据。网络拓扑结构如图2所示,此处引用三亚1扇60 min时间序列共1000个数据绘制可视图网络拓扑图,为方便研究将原拓扑图2(a)通过 Gephi 软件自动整理为图2(b)。



(a) Original network diagram (b) Organized topology diagram

图2 网络拓扑结构图

Fig.2 Network topology diagram

3.1 算法预测精度

实验过程中将训练集占数据集的比例定义为 proportion,根据 proportion 随机地划分为训练集与测试集。这个比例越大,表示给算法的信息越多,理论上算法的预测也越准确。当链路预测方法的准确度在0.5以下时,预测效果低于随机选取法,没有比较预测精确度的意义。因此,第一次实验 proportion 的值为0.9,其后每次实验减少0.1,直到0.5。

由于Katz、LHN-II和RWR中都有可调参数,为了更好地体现方法的预测效果,实验中对Katz分别取 $\beta = 0.01$ 和 $\beta = 0.0001$,对LHN-II取 $\phi = 0.9, \phi = 0.95, \phi = 0.99$,带有重启的随机游走方法的参数 $c = 0.85$,即游走粒子运动到某个节点上都有0.15的概率回到初始节点。以上取值均为查阅大量文献后发现使用频率最高的可调参数值。

为降低误差,所有数据均为计算100次精度后的平均值。表1~3均以三亚1扇60 min空中交通流量时间序列为数据进行分析的AUC精度比较,为便于观察预测精度大小,图3~5分别为其对应

表1 10种基于局部信息的相似性算法AUC精度比较

Table 1 Comparison of AUC accuracy of ten similarity algorithms based on local information

Proportion	0.9	0.8	0.7	0.6	0.5
CN	0.968 7	0.964 4	0.949 8	0.930 0	0.907 7
Salton	0.971 5	0.966 3	0.948 3	0.928 5	0.906 3
Jaccard	0.969 2	0.965 3	0.947 8	0.928 7	0.906 7
Sorenson	0.859 7	0.856 9	0.843 2	0.814 7	0.790 2
HPI	0.972 6	0.964 9	0.947 0	0.925 8	0.904 5
HDI	0.967 9	0.964 7	0.947 4	0.928 9	0.906 9
LHN	0.957 6	0.953 2	0.936 9	0.920 3	0.901 2
AA	0.975 8	0.970 7	0.955 0	0.934 9	0.912 0
RA	0.978 1	0.972 5	0.956 1	0.936 1	0.912 8
PA	0.675 7	0.691 4	0.722 6	0.707 8	0.712 6

表2 6种基于路径的相似性算法AUC精度比较

Table 2 Comparison of AUC accuracy of six path-based similarity algorithms

Proportion	0.9	0.8	0.7	0.6	0.5
Local Path	0.971 8	0.968 8	0.966 8	0.960 4	0.959 3
Katz0.01	0.970 7	0.968 6	0.967 4	0.966 3	0.965 7
Katz0.001	0.970 9	0.968 7	0.967 5	0.966 5	0.965 8
LHN II 0.9	0.957 7	0.953 6	0.946 2	0.946 6	0.944 8
LHN II 0.95	0.942 0	0.939 4	0.933 4	0.935 4	0.934 6
LHN II 0.99	0.894 6	0.893 5	0.893 8	0.896 1	0.897 2

表3 6种基于随机游走的相似性算法AUC精度比较

Table 3 Comparison of AUC accuracy of six similarity algorithms based on random walk

Proportion	0.9	0.8	0.7	0.6	0.5
ACT	0.622 4	0.669 2	0.670 5	0.670 9	0.692 9
RWR0.85	0.988 8	0.986 9	0.986 8	0.986 1	0.981 2
RWR0.9	0.987 4	0.985 3	0.985 2	0.985 3	0.980 8
RWR0.95	0.983 9	0.981 4	0.981 1	0.982 6	0.978 2
SRW0.85	0.988 8	0.986 9	0.986 9	0.986 1	0.981 2
LRW0.85	0.988 8	0.986 9	0.986 8	0.986 1	0.981 2

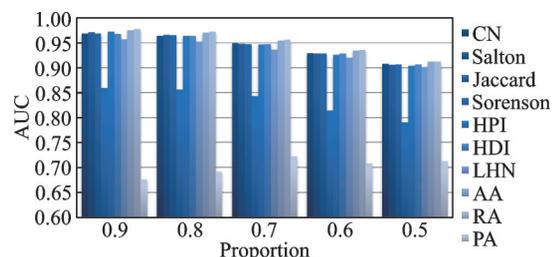


图3 不同比例的训练集在基于局部信息的算法中的AUC
Fig.3 AUC of different proportion training sets in local information-based algorithm

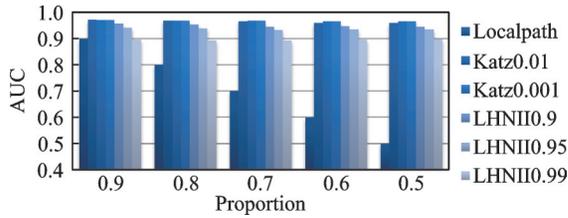


图 4 不同比例的训练集在基于路径的算法中的 AUC

Fig.4 AUC of different proportion training sets in path-based algorithm

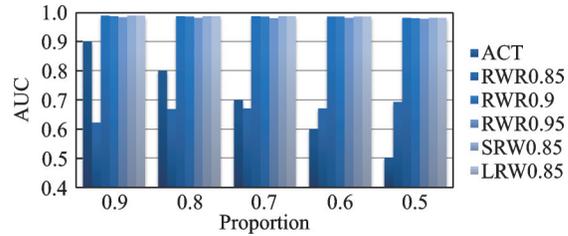


图 5 不同比例的训练集在基于随机游走的算法中的 AUC

Fig.5 AUC of training sets with different proportions in the algorithm based on random walk

的直方图。

由表 1~3 可以看出,绝大多数算法的预测精度都是随着 proportion 值的降低,AUC 值也随之降低。只有在基于路径的相似性算法中 LHN II 0.99 的预测效果是随着 proportion 值的降低,AUC 值逐渐增大,但增长幅度不大。其他大多数算法虽然降低幅度不大,但整体上仍以 0.9 为 proportion 值时预测效果最好。因此,为保证预测准确性,将 proportion 值设定为 0.9。

表 4 为三亚 1、2、4 三个扇区空中流量 60 min 时间尺度的各个链路预测方法 AUC 和 Precision 的精度值。在基于局部信息的 10 个算法中,RA 的预测精度较高,在 3 个扇区中都表现最好,均能够达到 97.7% 以上,其次是 AA,再次是 HPI。总体看来,PA 表现最差,特别是在三亚 2、4 扇区网络中已经低于 0.7。在基于路径的 6 个算法中,以 AUC 为

表 4 3 个扇区在不同算法中的 AUC 与 Precision 精度

Table 4 Comparison of AUC and Precision of three sectors in different algorithms

Proportion	三亚 1 扇 60 min		三亚 2 扇 60 min		三亚 4 扇 60 min	
	AUC	Precision	AUC	Precision	AUC	Precision
0.9						
CN	0.970 6	0.122 5	0.974 1	0.083 7	0.972 4	0.029 1
Salton	0.970 7	0.029 4	0.976 2	0.049 3	0.972 4	0.052 3
Jaccard	0.968 2	0.034 3	0.974 9	0.039 4	0.971 0	0.058 1
Sorenson	0.848 1	0.034 3	0.866 7	0.054 2	0.866 5	0.023 3
HPI	0.972 6	0.049	0.974 2	0.093 6	0.974 1	0.139 5
HDI	0.966 8	0.029 4	0.974 0	0.049 3	0.969 5	0.034 9
LHN	0.954 7	0.024 5	0.960 4	0.102 4	0.965 0	0.064 0
AA	0.976 9	0.112 7	0.982 0	0.073 9	0.976 7	0.064 0
RA	0.978 9	0.156 9	0.985 1	0.108 4	0.977 9	0.197 7
PA	0.709 8	0.063 7	0.685 5	0.098 5	0.643 0	0.017 4
Local Path	0.972 4	0.127 5	0.967 6	0.054 2	0.984 7	0.023 3
Katz0.01	0.971 5	0.142 2	0.965 4	0.059 1	0.984 6	0.046 5
Katz0.001	0.971 8	0.127 5	0.965 7	0.093 6	0.984 5	0.058 1
LHNII0.9	0.956 9	0.019 6	0.947 4	0.024 6	0.974 2	0.011 6
LHNII0.95	0.941 3	0.009 8	0.927 0	0.009 9	0.968 1	0.011 6
LHNII0.99	0.891 6	0.004 9	0.872 3	0.004 9	0.952 4	0.005 8
ACT	0.683 7	0.156 9	0.609 9	0.108 4	0.684 7	0.145 3
RWR0.85	0.989 9	0.186 3	0.987 9	0.128 1	0.992 1	0.168 6
RWR0.9	0.985 8	0.147 1	0.986 8	0.113 3	0.989 8	0.186 0
RWR0.95	0.983 4	0.107 8	0.981 4	0.103 4	0.988 3	0.174 4
SRW0.85	0.980 2	0.161 8	0.986 3	0.113 3	0.991 3	0.133 7
LRW0.85	0.985 7	0.117 6	0.983 7	0.098 5	0.991 5	0.157 0

参考,3 个扇区内 Local Path 表现最好,其次是 Katz。以 Precision 为参考,Katz0.01 预测精度更高。在基于随机游走的 6 个算法中,RWR0.85 预测效果最好,且高于基于局部信息的 RA 和基于路径的 Local Path,是针对 3 个扇区网络而言预测精度最高的链路预测方法。图 6 为以 0.9 比例划分的三亚 1 扇 60 min 时间序列在不同算法中的 AUC 与 Precision 精度值直方图对比。

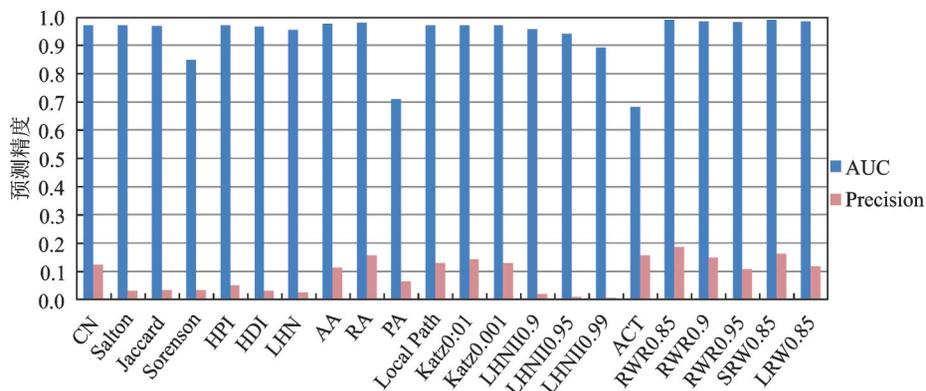


图 6 三亚 1 扇 60 min 训练集的 AUC 与 Precision 精度值

Fig.6 AUC and Precision values of a 60 min training set of Sanya 01 Sector

3.2 流量预测结果

选用三亚1扇2017年9月2日零时起24小时流量数据预测第25小时的流量值。首先将时间序列转换为可视图并根据可视性原则进行了标注,如图7所示。由3.1节中可以得知对于三亚扇区空中交通流量时间序列在不同种类的算法中RWR0.85的预测效果最好,因此选用此链路预测方法,将RWR算法参数 c 设置为0.85,即游走粒子运动以0.15的概率回到初始位置。

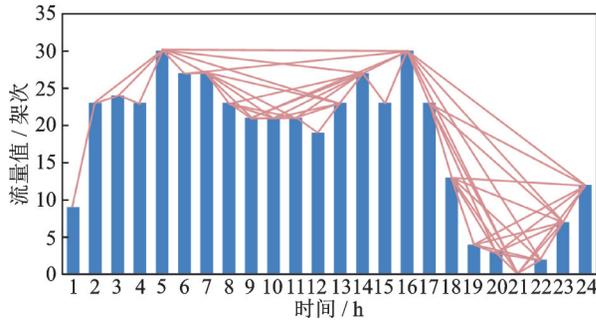


图7 三亚1扇9月2日24小时流量可视图

Fig. 7 24-hour traffic visualization map of Sanya 01 Sector on September 2nd

在执行两次随机游走后,得到节点相似度 S_{N1} 、 S_{N2} ,如表5所示,其中还标明了每个节点的节点度 k 。

图8为三亚1扇9月2日起60 min时间尺度的15天流量值直方图,时间序列整体趋势具有周期性,因此将 w 设置为4。通过观察得到重合的 w 个最大相似度时间节点依次为18、20、17,利用式(5)得到预测值依次为12、14、10。利用式(3)得到合成的相似度分别为0.3519、0.3413、0.3068。利用式(9)得到权重依次为0.5877、0.2017、0.2106。得到最终预测值 $y_{N+1}=0.5877 \times 12 + 0.2017 \times 14 + 0.2106 \times 10 = 11.9822$,而实际第25节点流量值为12,预测精度99.85%。其后依次滚动一个时间节点,得到预测流量值与原始流量值折线图,如图9所示。

三亚一扇60 min、30 min、15 min时间尺度未来6小时预测值如表6所示,可以看出随着时间序列划分越来越细,预测精度整体逐渐降低,且会出现在实际值存在突变时预测精度较低的情况。一方面,随着时间尺度变小流量值变化增大,为了保证预测的精度,需要更多的训练数据进行预测。另一方面也说明在实际航班运输过程中存在无法预测的意外情况,空中交通管制人员及其他相关人员应做好相应准备。

从表7中可以观察到,由于预测时最后已知节点 (t_N, y_N) 的影响最大,因此在出现短期数据大幅

表5 节点度和两次节点相似度的计算结果

Table 5 Calculation results of node degree and twice node similarity

t	S_{N1}	S_{N2}	流量值	k
1	0.014 2	0.014 0	9	2
2	0.023 8	0.023 2	23	3
3	0.025 2	0.023 6	24	3
4	0.024 2	0.025 8	23	2
5	0.064 0	0.062 2	30	6
6	0.063 4	0.058 2	27	3
7	0.069 2	0.067 6	27	8
8	0.035 2	0.062 8	23	7
9	0.059 2	0.061 8	21	5
10	0.040 2	0.040 0	21	4
11	0.036 6	0.039 2	21	4
12	0.020 8	0.024 4	19	2
13	0.045 4	0.047 2	23	2
14	0.070 0	0.069 0	27	3
15	0.069 8	0.066 2	23	2
16	0.268 8	0.248 2	30	7
17	0.298 2	0.313 0	23	8
18	0.327 6	0.326 8	13	7
19	0.216 2	0.312 8	4	5
20	0.315 8	0.328 8	3	5
21	0.222 8	0.296 4	0	4
22	0.213 8	0.242 6	2	2
23	0.309 2	0.237 8	7	2
24	0.912 2	0.913 4	12	1

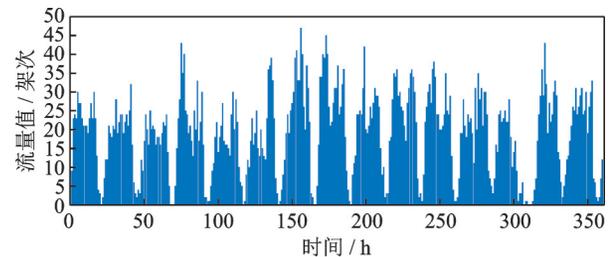


图8 三亚1扇9月2日起15天流量图

Fig.8 Flow value chart of Sanya 01 Sector for 15 days from September 2nd

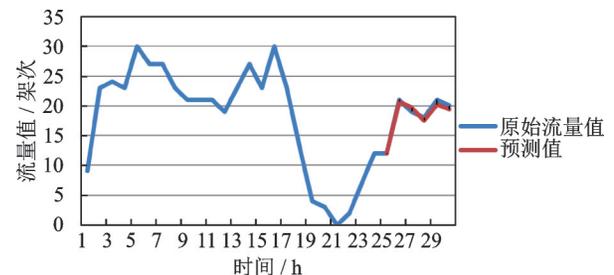


图9 三亚1扇6小时预测

Fig.9 6-hour forecast of Sanya 01 Sector

度波动时易有较大的预测偏差。且由于训练集中选取的时间段为0时起至某一天的24时,因此

表 6 使用 D-S 方法的预测值
Table 6 Predictive value with D-S method

时间	三亚 1 扇 60 min			三亚 1 扇 30 min			三亚 1 扇 15 min		
	原始流量值	预测值	预测精度	原始流量值	预测值	预测精度	原始流量值	预测值	预测精度
25	12	11.98	0.998 5	11	10.31	0.937 5	13	12.34	0.949 3
26	21	20.69	0.985 2	13	12.49	0.960 5	9	9.77	0.921 1
27	19	18.35	0.965 9	20	19.48	0.973 8	12	11.31	0.942 3
28	18	17.54	0.974 6	13	13.57	0.957 9	8	8.58	0.932 2
29	21	20.24	0.963 8	16	15.44	0.965 0	3	8.94	0.335 4
30	20	19.35	0.967 6	9	10.28	0.875 2	8	7.30	0.912 3
31				16	15.53	0.970 6	11	10.16	0.923 9
32				21	20.31	0.967 1	13	12.32	0.947 8
33				15	15.79	0.950 2	8	8.76	0.913 5
34				13	12.60	0.969 2	7	6.27	0.895 9
35				12	11.58	0.965 0	8	7.68	0.960 5
36				3	9.27	0.323 7	8	7.74	0.967 8
37							6	7.00	0.856 7
38							9	8.21	0.912 0
39							10	9.45	0.945 2
40							10	9.38	0.937 9
41							11	10.47	0.952 1
42							7	9.79	0.715 3
43							5	7.32	0.682 9
44							6	6.32	0.949 2
45							10	7.40	0.740 2
46							7	7.90	0.885 7
47							7	7.34	0.954 1
48							6	7.03	0.854 0

表 7 未使用 D-S 方法的预测值
Table 7 Predictive values without D-S method

时间	三亚 1 扇 60 min			三亚 1 扇 30 min			三亚 1 扇 15 min		
	原始流量值	预测值	预测精度	原始流量值	预测值	预测精度	原始流量值	预测值	预测精度
25	12	11.59	0.965 8	11	10.37	0.942 7	13	12.33	0.948 5
26	21	13.59	0.647 1	13	12.12	0.932 3	9	10.29	0.874 6
27	19	17.66	0.929 5	20	19.28	0.964 0	12	11.09	0.924 2
28	18	18.45	0.975 6	13	14.01	0.927 9	8	10.83	0.738 7
29	21	20.18	0.961 0	16	15.27	0.954 4	3	7.52	0.398 9
30	20	19.66	0.983 0	9	12.93	0.696 1	8	6.55	0.818 8
31				16	10.44	0.652 5	11	10.13	0.920 9
32				21	18.99	0.904 3	13	11.96	0.920 0
33				15	19.58	0.766 1	8	12.48	0.641 0
34				1	15.60	0.064 1	7	7.11	0.984 5
35				12	6.89	0.574 2	8	7.96	0.995 0
36				3	5.37	0.558 7	8	8.01	0.998 8
37							6	8.00	0.750 0
38							9	7.60	0.844 2
39							10	9.34	0.934 1
40							10	9.94	0.993 7
41							11	10.01	0.909 6
42							7	10.83	0.646 1
43							5	9.12	0.548 2
44							6	6.33	0.947 9
45							10	7.86	0.786 0
46							7	8.29	0.844 0
47							7	7.31	0.958 1
48							6	6.84	0.877 3

60 min 时间尺度预测的时间节点,为未来一天的 0 时至 6 时,必然会存在与昼间流量形成骤减状态的时刻。在如表 7 中三亚 1 扇 60 min 所在列,26 时间所在行可以看出原始流量值由 12 剧增至 21,极大地影响了预测效果,与 25 时 0.965 8 的预测精度相比 26 时预测精度只有 0.647 1。而未使用 D-S 的预测精度均远低于如表 6 所示的使用 D-S 方法的预测精度。可以证明 D-S 方法的使用是可行且有效的。

4 结 论

基于复杂网络链路预测理论对三亚区域管制扇区的空中交通流量进行了基于局部信息、路径和随机游走的共 22 种算法精度比较和流量值预测。由于链路预测的方法只能预测连边,而不能预测节点,不能直接预测出流量值。因此,需要在链路预测的基础上,针对关键节点的权重问题引入介绍了 D-S 证据理论。最后列出了流量值预测的步骤方法。主要结论如下:

(1) 基于随机游走的 RWR0.85 算法对预测三亚扇区空中交通流量精度最高。在基于局部信息的算法中 RA 预测精度最高,基于路径的算法中 Local Path 预测精度最高。

(2) 在经过流量预测后还发现未使用 D-S 的预测精度均远低于使用 D-S 方法的预测精度。可以证明 D-S 方法的使用是可行且有效的。

(3) 随着时间序列划分越来越细,预测精度整体逐渐降低,出现在实际流量值存在突变时预测精度较低的情况。说明在实际航班运输过程中存在无法预测的意外情况,空中交通管制人员及其他相关人员应做好相应准备。

(4) 应用复杂网络链路预测及 D-S 证据理论方法来预测空中交通流量是可行有效的,为进一步研究和提高预测精度奠定了基础。

参考文献:

- [1] 葛柏君. 短期区域飞行流量预测问题研究[D]. 南京: 南京航空航天大学, 2008.
GE Baijun. Research on short-term regional flight flow forecast[D]. Nanjing: Nanjing University of Aeronautics and Astronautics, 2008.
- [2] GETOOR L, DIEHL C P. Link mining: A survey [J]. ACM SIGKDD Explorations Newsletter, 2005, 7(2): 1-10.
- [3] 周磊, 龚志强, 支蓉, 等. 利用复杂网络研究中国温度序列的拓扑性质[J]. 物理学报, 2008(11): 7380-7389.
ZHOU Lei, GONG Zhiqiang, ZHI Rong, et al. Topological properties of China temperature series are studied by using complex networks[J]. Acta Physica Sinica, 2008(11): 7380-7389.
- [4] O' MADADHAIN J, HUTCHINS J, SMYTH P. Prediction and ranking algorithms for event-based network data[J]. ACM SIGKDD Explorations Newsletter, 2005, 7(2): 23-30.
- [5] LI Xin, CHEN Hsinchun. Recommendation as link prediction in bipartite graphs: A graph kernel-based machine learning approach[J]. Decision Support Systems, 2013, 54(2): 880-890.
- [6] 吕琳媛. 复杂网络链路预测[J]. 电子科技大学学报, 2010, 39(5): 651-661.
LV Linyuan. Complex network link prediction[J]. Journal of University of Electronic Science and Technology of China, 2010, 39(5): 651-661.
- [7] 王超, 朱明, 赵元棣. 基于改进加权一阶局域法的空中交通流量预测模型[J]. 西南交通大学学报, 2018, 53(1): 206-213.
WANG Chao, ZHU Ming, ZHAO Yuandi. Air traffic flow prediction model based on improved weighted first-order local area method[J]. Journal of Southwest Jiaotong University, 2018, 53(1): 206-213.
- [8] 宋维凯, 张洪海, 万俊强, 等. 基于动态时间规整和长短期记忆的空中交通流量短期预测[J]. 科学技术与工程, 2022, 22(25): 11256-11261.
SONG Weikai, ZHANG Honghai, WAN Junqiang, et al. Short term prediction of air traffic flow based on dynamic time warping and short-term memory[J]. Science, Technology and Engineering, 2022, 22(25): 11256-11261.
- [9] 韩昀轩. 基于交通分配的航路网络生成[D]. 南京: 南京航空航天大学, 2019.
HAN Yunxuan. Route network generation based on traffic assignment[D]. Nanjing: Nanjing University of Aeronautics and Astronautics, 2019.
- [10] 曹希. 基于复杂网络理论的航路网络生成及优化[D]. 南京: 南京航空航天大学, 2017.
CAO Xi. Route network generation and optimization based on complex network theory[D]. Nanjing: Nanjing University of Aeronautics and Astronautics, 2017.
- [11] CLAUSET A, MOORE C, NEWMAN M E J. Hierarchical structure and the prediction of missing links in networks[J]. Nature, 2008, 453: 98-101.
- [12] LIBEN-NOWELL D, KLEINBERG J O N. The link-prediction problem for social networks[J]. Journal of the American Society for Information Science and Technology, 2007, 58: 1019-1031.
- [13] LV Linyuan, ZHOU Tao. Link prediction in complex networks: A survey[J]. Physica A: Statistical Mechanics and Its Applications, 2011, 390(6): 1150-1170.

- [14] 赵延乐. 基于复杂网络的链路预测研究[D]. 保定: 河北大学, 2014.
ZHAO Yanle. Research on link prediction based on complex networks[D]. Baoding: Hebei University, 2014.
- [15] 吕亚楠. 基于网络结构和随机游走理论的链路预测算法研究[D]. 武汉: 武汉理工大学, 2019.
LV Yanan. Research on link prediction algorithm based on network structure and random walk theory [D]. Wuhan: Wuhan University of Technology, 2019.
- [16] RAVASZ E, SOMERA A L, MONGRU D A, et al. Hierarchical organization of modularity in metabolic networks[J]. Science, 2002, 297(5586): 1553-1555.
- [17] LEICHT E A, HOLME P, NEWMAN M E J. Vertex similarity in networks[J]. Physical Review E, 2006, 73(2): 1-10.
- [18] ADAMIC L, ADAR E. Friends and neighbors on the web[J]. Social Networks, 2003, 25(3): 211-230.
- [19] ZHOU Tao, LV Linyuan, ZHANG Yicheng. Predicting missing links via local information[J]. The European Physical Journal, B: Condensed Matter Physics, 2009, 71(4): 623-630.
- [20] LV Linyuan, HANG Jinci, ZHOU Tao. Similarity index based on local paths for link prediction of complex networks[J]. Physical Review E: Statistical, Nonlinear, and Soft Matter Physics, 2009, 80(4): 046122.
- [21] LIU Weiping, LV Linyuan. Link prediction based on local random walk[J]. Europhysics Letters, 2010, 89(5): 1-6.
- [22] 饶君, 吴斌, 东昱晓. MapReduce 环境下的并行复杂网络链路预测[J]. 软件学报, 2012, 23(12): 3175-3186.
RAO Jun, WU Bin, DONG Yuxiao. Link prediction of parallel complex networks in MapReduce environment [J]. Journal of Software, 2012, 23(12): 3175-3186.
- [23] HERLOCKER J L, KONSTAN J A, TERVEEN L G, et al. Evaluating collaborative filtering recommender systems[J]. ACM Transactions on Information Systems, 2004, 22(1): 5-53.
- [24] 邱望仁, 刘晓东. 基于证据理论的模糊时间序列预测模型[J]. 控制与决策, 2012, 27(1): 99-103.
QIU Wangren, LIU Xiaodong. Fuzzy time series prediction model based on evidence theory[J]. Control and Decision, 2012, 27(1): 99-103.
- [25] 李霞, 李守伟. 基于EMD与DVG的非线性时间序列预测模型及其应用研究[J]. 中国管理科学, 2022, 30(9): 275-286.
LI Xia, LI Shouwei. Research on nonlinear time series prediction model based on EMD and DVG and its application[J]. China Management Science, 2022, 30(9): 275-286.
- [26] MAO Shengzhong, XIAO Fuyuan. Time series forecasting based on complex network analysis[J]. IEEE Access, 2019, 7: 40220-40229.
- [27] ZHANG Rong, AAHURI Baabak, SHYR Yu, et al. Forecasting construction cost index based on visibility graph: A network approach[J]. Physica A: Statistical Mechanics and its Applications, 2018, 493: 239-252.

(编辑:夏道家)