大型多人在线角色扮演游戏的下一地点预测

佟振明[1] 刘志鹏[2]

(三江学院计算机科学与工程学院 南京 210012)[1]

(南京邮电大学现代邮政学院、现代邮政研究院 南京 210003)[2]

摘 要 近年来,大型多人在线角色扮演游戏(MMORPG)已经成为最流行的网络娱乐活动之一。MMORPG 在游戏环境中形成虚拟社会,其中每个玩家扮演某个虚构角色,并控制该角色的大多数活动。游戏的迅猛发展累积了海量数据,其中包含游戏虚拟社会的语义和拓扑信息。研究者针对游戏数据开展了一系列研究工作,如玩家退出预测、游戏服务器整合等。游戏角色的下一地点预测对提升游戏体验、改善游戏设计和检测游戏机器人均有十分重要的意义。目前,该项预测任务主要使用统计分析完成。然而,由于游戏数据具有海量特征,因此需要一种自动化的计算方法。文中提出了基于隐马尔科夫模型的游戏角色下一地点预测模型,该模型能够考虑与位置特性相关的不可观测的属性,同时兼顾游戏角色前期行为的影响。实验结果表明,与现有方法相比,该方法具有建模直观的特点,在稠密分布的 MMORPG 数据中能够得到更准确的下一地点预测结果。

关键词 下一位置预测,大型多人在线角色扮演游戏,隐马尔可夫模型,游戏日志挖掘

中图法分类号 TP391 **文献标识码** A

Next Place Prediction of Massively Multiplayer Online Role-playing Games

TONG Zhen-ming[1] LIU Zhi-peng[2]

(College of Computer Science and Engineering, Sanjiang University, Nanjing 210012, China)[1]

(School of Modern Posts and Institute of Modern Posts, Nanjing University of Posts and Telecommunications, Nanjing 210003, China)[2]

Abstract In recent years, massively multiplayer online role-playing games (MMORPG) has become one of the most popular Internet recreational activities. MMORPG creates virtual societies, in which each user plays a fictional character, and controls most of its activities. With rapid development of MMORPG, it has accumulated massive data, which contain semantic as well as topological information of virtual societies. Researchers have already carried out many studies, such as player departure prediction and server consolidation. The task of next place prediction is crucial to enhance gaming experience, improve game design and game bot detection, and most of next place prediction methods are based on statistical analysis. However, it is difficult to apply these methods in practice due to the characteristic of large scale of game data, and an automatic computation method to be developed. This paper proposed a next place prediction algorithm based on hidden Markov model (HMM). The model considers location characteristics as unobservable parameters, and takes the effects of previous actions of each game character into consideration. Experimental results with real MMORPG dataset show that our approach is intuitive and has better performance in dense distributed data than other existing methods for the task of next place prediction of MMORPG.

Keywords Next place prediction, MMORPG, Hidden Markov model, Game data mining

1 引言

近年来,大型多人在线角色扮演游戏(MMORPG)已经成为最流行的网络娱乐活动之一。MMORPG 的成功范例之一是魔兽世界,该游戏中有多个种族和多个角色,在网络空间中形成了多个虚拟社会。其中,每个游戏玩家扮演某个虚构角色,并控制该角色的大多数活动。玩家之间形成了复杂的虚拟社会关系和团体,这些社会关系包括朋友、决斗和交易等。社会团体包括为了达到某个目的形成的短期团体和长期稳定的行会组织等。这些社会关系和团体产生了大量的虚拟社会活动[1]。MMORPG 的迅猛发展累积了海量游戏数据,以魔兽世界数据集为例[2],每一项数据包含查询时间、查询序列号、游戏 ID 号、该游戏 ID 所属的行会、游戏等级、种族以及执行查询时虚拟游戏人物所在区域等信息。这些游戏数据包含游戏虚拟社会的语义和拓扑信息。针对游戏数据开展挖掘工作有重大的经济价值。目前已经开展的典型研究工作如玩家退出预测[3]、游戏服务器整合[4]和游戏机器人检测[5-6]等。文献[7-8]使用统计学方法对整个游戏虚拟社会的种族分布、虚拟游戏人物的性别比例、游戏等级进化和游戏团体演变等问题展开研究。这些工作均基于统计学方法,并由人工分析提

本文受南京邮电大学校级科研基金(NY214126)资助。

佟振明(1963—),男,副教授,主要研究领域为算法与数据结构,E-mail:qtdscn@163.com;**刘志鹏**(1980—),男,博士,副教授,主要研究领域为社交网络数据挖掘,E-mail:liuzhipengcs@139.com。

取数据的统计特性，然后利用该统计特性开展预测或检测工作。但 MMORPG 的数据具有海量性特征，因此开发自动化的数据挖掘算法具有紧迫性和必要性。

魔兽世界等 MMORPG 虚拟游戏角色的下一地点预测，对提升游戏体验、改善游戏设计和检测游戏机器人均有一定的意义：1)在线游戏场景切换需要加载大量数据，这需要一段游戏加载时间，如果能够较准确地预测游戏角色的下一地点，则可以提前加载下一场景的部分游戏数据，缩短场景切换时的时间代价，从而在一定程度上提升游戏体验；2)通过分析虚拟游戏角色的移动特征，可以了解游戏人物更倾向完成哪些游戏场景的任务，哪些游戏场景设置得过难或过于简单，进而通过改善现有游戏场景或设置新的游戏场景来改进游戏设计；3)与人工控制相比，游戏机器人控制的虚拟游戏人物的运动更有规律[6,9-10]。通过下一地点预测，可以筛选出具有高度运动规律的虚拟游戏人物，为游戏机器人的检测工作提供服务。

当前，已经有大量针对 GPS 位置数据的下一地点预测算法。诺基亚举办的移动数据挖掘竞赛中，挖掘任务之一便是下一地点预测[11-13]。文献[11]针对单个用户的运动轨迹，使用动态贝叶斯网络、神经网络和梯度提升决策树 3 种预测模型，并采用 5 种组合策略来提升算法的预测准确性。文献[12]使用周期性模型和多类分类策略，最终采用了基于周期和 SVM 的 6 种算法开展预测工作，其准确率最高能够达到 55.69%。文献[13]利用时空上下文信息，提出了 HPHD 模型，并将其与 9 种基线算法进行比较，其预测准确率达到 50.53%。文献[14]提取不同的移动数据时空特性，提出了 18 种预测下一地点的策略。文中提出的算法 MAJOR 使用 18 种策略的预测结果，运用启发式的多数投票算法进行下一地点的预测。文献[15]使用复杂的数据预处理策略，这些数据的处理步骤由家庭和工作场所检测、假期检测、新地点检测等组成；然后使用决策树算法进行下一位置的预测。文献[16]综合考虑用户在不同类型位置之间的过渡信息、位置场景之间的移动性和用户登录系统模式的时空特性等信息，使用两种基于线性回归和 M5 模型树算法在 Foursquare 数据集上开展预测工作。实验结果表明，M5 模型树预测算法的性能较好。上述工作都需要从 GPS 坐标信息中识别出有意义的位置信息，如工作场所和家等，并在此基础上展开研究。由于移动数据包含的数据类型和数据特征较多，这些方法由人工提取和融合多种数据特征、复杂的数据预处理技术以及组合运用多种数据预测模型完成，因此算法的预测准确率有限，很多算法的预测准确率不到 60%。

2 数据建模

2.1 前提

本文提出的下一地点预测算法基于以下 3 个前提[7]：1)在线游戏的场景设置各不相同，由于各个场景内资源分配不均等，75%以上的用户集中在 18%的场景内活动；2)在线游戏虚拟人物的行动轨迹可以表示为一系列游戏用户的请求序列；3)人类操作的游戏行为具有一定的随机性，游戏机器人难以模拟真实用户的游戏行为。游戏机器人的功能往往重复执行某项特定功能来提升等级。魔兽世界中可以使用 Lua 编写的脚本反复执行钓鱼任务，以提升游戏任务的钓鱼技能。与真实用户的游戏行为相比，游戏机器人产生的轨迹序列具有周期性和高度可预测性。

2.2 建模

本文实验使用魔兽世界的历史数据[2]，数据采集自台湾的圣光之愿(Light's Hope)服务器。数据采集程序使用 Lua 脚本编写，该脚本的功能类似快照，使用 who 命令查看并保存当前服务器上的在线游戏人物的基本信息。该脚本的运行周期为 10 分钟，如果在两次脚本运行期间有某个用户上线并下线，则历史数据无法保存该用户的活动信息。数据采集自 2006 年 1 月到 2009 年 1 月，共持续 1107 天。采集游戏人物共 91065 人。由于服务器维护和客户端兼容性等问题，共采集了 138084 个系统快照，缺失 21324 个快照信息。快照中的每个数据项包含如下基本信息：查询时间、查询序列号、游戏人物的 ID 号、所在工会、等级、种族、职业以及游戏人物所处的区域。这里的区域指代虚拟游戏世界的某个范围，其数值为魔兽世界中 229 个区域之一。大型多人在线角色扮演游戏的下一地点预测的任务，便是预测游戏人物在虚拟游戏世界中的下一个可能区域。

魔兽世界历史数据的建模过程如图 1 所示。游戏用户在客户端登录游戏，操纵角色执行游戏任务，且在执行任务的过程中与服务器产生频繁的数据交互。服务器周期性运行脚本采集数据，获取游戏角色的相关信息，如所在区域等。数据采集完成后，进入数据分析环节。首先是基于数据的区域识别。如果游戏角色没有切换所在的游戏区域，则脚本得到的区域数值不变；如果在此期间游戏角色切换游戏区域，则脚本得到的区域数值发生变化。据此，可以从采集的数据中获取用户切换游戏区域的时间区间，根据区域识别结果构建隐马尔可夫模型，从而进行下一地点预测。

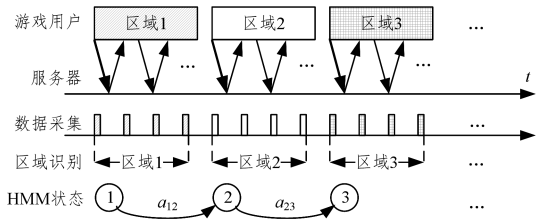


图 1 魔兽世界的下一地点预测建模

3 隐马尔可夫模型

隐马尔可夫模型(Hidden Markov Model，HMM)描述一个含有隐含未知参数的马尔可夫过程，无法观测其系统状态[17]。当到达某个状态时，系统记录观测信息，该观测是该状态的一个概率函数。假设系统任意时刻处于 N 个离散状态之一：$S_1,S_2,\cdots,S_N$；时刻 t 的状态为 q_t，$t=1,2,\cdots$；Q 为状态序列；每个状态的离散观测 $v_m\in\{v_1,v_2,\cdots,v_M\}$，$M$ 为系统可观测状态数；$b_j(m)$ 为系统在状态 S_j 时观测到 v_m ($m=1,\cdots,M$)的概率；O 为观测序列。假定观测概率不依赖于时间 t，则可由一系列观测状态形成 O。此时，状态序列 Q 是不可观测的，仅能通过观测序列推断。

给定 HMM 模型 $\lambda=(A,B,\Pi)$，系统由如下 5 个部分组成。

1)N：模型状态个数，$S=\{S_1,S_2,\cdots,S_N\}$；

2)M：按序排列的不同观测符号的个数，$V=\{v_1,v_2,\cdots,v_M\}$；

3）状态转移概率，$A=[a_{ij}]$，$a_{ij}=P(q_{t+1}=S_j|q_t=S_j)$；

4）观测概率，$B=b_j(m)$，$b_j(m)=P(O_t=v_m|q_t=S_j)$；

5）初始状态概率，$\Pi=[\pi_i]$，$\pi_i=P(q_1=S_i)$。

3.1 HMM 模型参数学习

给定一系列观测序列组成的训练集 $\chi=\{O^k\}_k^\omega$，$\omega=1,2,\cdots,K$。假定这些观测序列相互独立，即 $P(\chi|\lambda)=\prod_{k=1}^{K}P(O^k|\lambda)$，HMM 模型参数学习过程采用最大似然方法，计算最大化 $P(\chi|\lambda)$ 的 λ^*。

假设 $\xi_t(i,j)$ 为给定 O 和 λ，在时刻 t 处于 S_i，且在时刻 $t+1$ 处于 S_j 的概率。$\alpha_t(i)$ 为正向变量，表示时刻 t 之前 t 个观测 $\{O_1,O_2,\cdots,O_t\}$ 终止于 S_i 的概率，$\alpha_t(i)=P(O_1,\cdots,O_t,q_t=S_i|\lambda)$。类似地，$\beta_{t+1}(j)$ 为反向变量，表示时刻 $t+1$ 处于状态 S_j 且观测到部分序列 $O_{t+2},\cdots,O_T$ 的概率，即 $\beta_{t+1}(j)=P(O_{t+2},\cdots,O_T|q_{t+1}=S_j,\lambda)$。$\alpha_t(i)$ 以概率 a_{ij} 转移到 S_j，产生 $\beta_{t+1}(j)$。则：

$$\xi_t(i,j)=P(q_t=S_i,q_t=S_j|O,\lambda)=\frac{\alpha_t(i)a_{ij}b_j(O_{t+1})\beta_{t+1}(j)}{\sum_k\sum_l\alpha_t(k)a_{kl}b_l(O_{t+1})\beta_{t+1}(l)}$$

假设时刻 t 系统处于状态 S_i 的概率为 $\gamma_t(i)$，则 $\gamma_t(i)=\sum_{j=1}^{N}\xi_t(i,j)$。

HMM 模型参数学习使用基于 EM 方法的 Baum-Welch 算法。假设 z_i^t 代表 $q_t=S_i$ 的概率，z_{ij}^t 代表 $q_t=S_i$ 且 $q_{t+1}=S_j$ 的概率。该算法基于迭代执行如下两个步骤，直至计算结果收敛。

1）给定 $\lambda=(A,B,\Pi)$，计算 $\xi_t(i,j)$ 和 $\gamma_t(i)$，则有：

$E[z_i^t]=\gamma_t(i)$

$E[z_{ij}^t]=\xi_t(i,j)$

2）给定 $\xi_t(i,j)$ 和 $\gamma_t(i)$，计算 λ。系统参数在所有观测上求平均值：

$$\hat{a}_{ij}=\frac{\sum_{k=1}^{K}\sum_{t=1}^{T_K-1}\xi_t^k(i,j)}{\sum_{k=1}^{K}\sum_{t=1}^{T_K-1}\gamma_t^k(i)}$$

$$\hat{b}_j(m)=\frac{\sum_{k=1}^{K}\sum_{t=1}^{T_K}\gamma_t^k(j)\cdot 1(O_t^k=v_m)}{\sum_{k=1}^{K}\sum_{t=1}^{T_K-1}\gamma_t^k(j)}$$

$$\hat{\pi}_i=\frac{\sum_{k=1}^{K}\gamma_1^k(i)}{K}$$

3.2 概率估计

当完成 HMM 模型参数的学习后，采用正反向过程计算任意给定观测序列 $O=\{O_1,O_2,\cdots,O_T\}$ 的概率 $P(O|\lambda)$。

1）计算 $\alpha_t(i)$ 的递推过程

初始化：$\alpha_1(i)=P(O_1,q_1=S_i|\lambda)=\pi_ib_i(O_1)$

递推：$\alpha_{t+1}(j)=P(O_1,\cdots,O_{t+1},q_{t+1}=S_j|\lambda)=[\sum_{i=1}^{N}\alpha_t(i)a_{ij}]b_j(O_{t+1})$

观测序列的概率：

$$P(O|\lambda)=\sum_{i=1}^{N}P(O,q_T=S_j|\lambda)=\sum_{i=1}^{N}\alpha_T(i)$$

2）计算 $\beta_t(i)$ 的递推过程

初始化：$\beta_T(i)=1$

递推：$\beta_t(i)=P(O_{t+1},\cdots,O_T|q_t=S_i,\lambda)=\sum_{j=1}^{N}a_{ij}b_jO(t+1)\beta_{t+1}(j)$

4 实验结果与分析

4.1 实验设置

本文采用魔兽世界的历史数据[2]进行实验分析。实验在 Pentium(R) D 3.0 GHz 的 PC 机器上进行，主存为 4 GB，运行 CentOS 4.5 操作系统。算法基于 R 语言的 HiddenMarkov[1)] 开源包实现。

实验主要将基于 HMM 的下一地点预测模型与下列计算方法进行比较。

1）决策树算法（Decision Tree，DT）[15]。该算法的思想是使用当前预测的游戏角色的数据特征构建决策树。这些特征主要包括系统时间以及游戏人物所处的区域。严格来讲，影响游戏人物下一地点的因素比数据集中的特征更为复杂，但该算法的特征提取受到当前数据集内容的限制。

2）动态贝叶斯网络（Dynamical Bayesian Network，DBN）[11]。构建动态贝叶斯网络的思想是：游戏人物的下一地点预测依赖于其当前位置，且同时赖于访问下一位置的起始时间。与下一位置访问时间间隔相近时，当前位置信息具有较强的参考价值。当两次访问时间间隔较大时，下一次访问的起始时间至关重要。

实验选择了魔兽世界游戏 3 年的历史数据来评估下一地点预测算法。实验中使用准确率 P_a 评估算法的有效性。P_a 是正确预测的下一位置数量 P_c 与总预测数量 P_t 的比值，即 $P_a=P_c/P_t$。

4.2 数据预处理

魔兽世界历史数据的预处理由两个重要步骤组成，分别是数据重组、数据分类。数据重组的原理如图 2 所示。原始魔兽世界的历史数据作为输入，按照时间顺序 t_i，用快照方式存储。数据重组过程从中抽取每个用户 U_j 在对应时间点 t_i 的位置信息 P_k，并将其存放到该用户 U_j 对应的文件中；数据重组后，每个游戏用户 U_j 文件内的位置信息按照时间顺序排列。

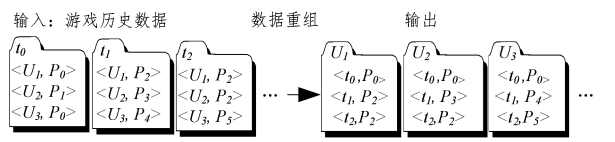


图 2　魔兽世界历史数据的预处理

数据分类过程将稀疏和稠密的用户数据分开。由于服务器维护和客户端兼容性等问题，系统缺失一部分用户数据。除此以外，还有两个造成用户数据稀疏的重要原因：1）游戏用户线上游戏活动不频繁，造成数据采集过于稀疏；2）用户在一段时间后退出游戏。用户数据稀疏导致该用户的游戏行为难以分析和预测。本文采用的数据预处理过程将游戏总时间不超过 600 天（即游戏记录数量不足 86400 个），或两次游戏时间间隔超过 5 天（即连续缺失 720 个数据记录）的游戏用户数据称为稠密组；其余的数据则称为稀疏组。这里的时间间隔

1) http://cran.r-project.org/web/packages/HiddenMarkov/index.html

是用户自定义的，每个用户可以根据自己的时间设定数据稀疏和稠密的临界值。

4.3 实验结果

图 3 给出了 3 种算法在稠密组数据上的总体平均预测准确率。其中，DT 的总体平均预测准确率为 55.27%，DBN 的总体平均预测准确率为 62.32%，HMM 的总体平均预测准确率为 71.30%，但这并不意味着 HMM 的下一地点预测算法一定比 DT 和 DBN 优秀。此处，DT 和 DBN 的预测性能在很大程度上受限于数据特征的选择。由于系统给定的魔兽世界历史数据集本身包含的数据特征不够丰富，因此在构建预测分类算法时可采用的数据特征显得不足。实验中发现，针对某个特定游戏用户的下一地点预测，DBN 的预测准确率有可能高于 HMM，这是因为 DBN 在预测时可以考虑如时间信息等更多的信息。HMM 下一地点预测算法对不同的用户而言准确率差别较大，部分游戏角色下一地点预测的准确率高达 85.7%～89.3%。出现此情况有两种可能：1)该用户偏爱某些游戏场景，因此反复执行该任务；2)该游戏用户使用了游戏机器人，这为检测游戏机器人提供了新的思路。由于数据信息不足，这里无法使用基于统计学的游戏机器人检测方法[5-6,9]对比结果。

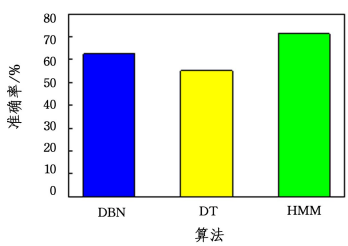


图 3 稠密组数据的总体平均预测准确率

图 4 给出了 3 种算法在稀疏组数据上的总体平均预测准确率。其中，DT 的总体平均预测准确率为 40.22%，HMM 的总体平均预测准确率为 46.31%，DBN 的总体平均预测准确率为 55.52%。与稠密组数据的预测准确率相比，各个算法在稀疏组数据上的预测率均有所下降。此时 HMM 的总体平均预测准确率比 DBN 低，这是由于 DBN 考虑了下一次访问位置的时间，在数据稀疏的情况下，该条件变得十分重要。

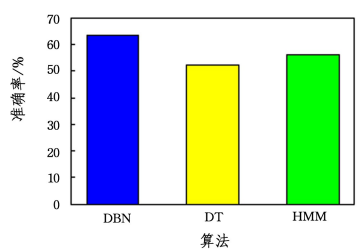


图 4 稀疏组数据的总体平均预测准确率

4.4 讨论

本文采用 HMM 进行游戏角色下一地点预测时存在一些局限性：

1)受数据集采集的特征种类的限制，无法提取更多有用的特征以提升 DBN 和 DT 算法的性能。

2)受数据集位置信息精度的限制，如果游戏角色的活动范围始终在一个区域内，没有产生跨区域的操作，则算法无法检测；如果有游戏角色位置数据，则可以采用三角网格等方法划分虚拟空间，进而采用 HMM 算法进行下一地点的预测。

结束语 本文研究了基于魔兽世界历史数据的下一地点预测问题，提出了基于 HMM 的下一地点预测模型。该过程由以下 3 个主要步骤组成：数据建模、数据预处理和 HMM 模型学习。该模型主要利用魔兽世界历史数据的位置信息，同时考虑前后位置之间的关联性，采用 HMM 模型预测下一地点。与传统方法相比，HMM 的下一地点预测模型具有建模直观且简单的重要特点，在数据较稠密的情况下获得了更高的预测率。

未来将从如下几个方面继续开展相关的研究工作：1)根据实验结果，在进行下一地点预测时，根据数据的实际情况综合选用多种算法，以提升算法准确率；2)本文中定义稀疏和稠密数据集的临界数值由人工指定，可以考虑采用统计学或机器学习算法，自动判定该临界数值；3)采用并行算法来提升算法的效率和性能。

参 考 文 献

[1] NARDI B, HARRIS J. Strangers and friends: Collaborative play in World of Warcraft[C]// International Handbook of Internet Research. Springer, 2010: 395-410.

[2] LEE Y T, CHEN K T, CHENG Y M, et al. World of Warcraft avatar history dataset[C]// Proceedings of the Second Annual ACM Conference on Multimedia Systems. ACM, 2011: 123-128.

[3] TARNG P Y, CHEN K T, HUANG P. On prophesying online gamer departure[C]// 2009 8th Annual Workshop on Network and Systems Support for Games (NetGames). IEEE, 2009: 1-2.

[4] LEE Y T, CHEN K T. Is server consolidation beneficial to MMORPG? A case study of World of Warcraft[C]// 2010 IEEE 3rd International Conference on Cloud Computing (CLOUD). IEEE, 2010: 435-442.

[5] CHEN K T, PAO H K K, CHANG H C. Game bot identification based on manifold learning[C]// Proceedings of the 7th ACM SIGCOMM Workshop on Network and System Support for Games. ACM, 2008: 21-26.

[6] CHEN K T, LIAO A, PAO H K K, et al. Game bot detection based on avatar trajectory[C]// Entertainment Computing-ICEC 2008. Springer, 2009: 94-105.

[7] DUCHENEAUT N, YEE N, NICKELL E, et al. The life and death of online gaming communities: a look at guilds in world of warcraft[C]// Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, 2007: 839-848.

[8] DUCHENEAUT N, YEE N, NICKELL E, et al. Building an MMO with mass appeal a look at gameplay in world of warcraft [J]. Games and Culture, 2006, 1(4): 281-317.

[9] MITTERHOFER S, PLATZER C, KRUEGEL C, et al. Server-side bot detection in massive multiplayer online games[J]. IEEE Security and Privacy, 2009, 7(3): 29-36.

[10] PLATZER C. Sequence-based bot detection in massive multiplayer online games[C]// 2011 8th International Conference on Information, Communications and Signal Processing (ICICS). IEEE, 2011: 1-5.

[11] ETTER V, KAFSI M, KAZEMI E. Been there, done that: What your mobility traces reveal about your behavior[C]// Mobile Data Challenge by Nokia Workshop, in Conjunction with Int

Conf on Pervasive Computing. 2012.

[12] WANG J, PRABHALA B. Periodicity based next place prediction[C]// Nokia Mobile Data Challenge 2012 Workshop Dedicated Task. Citeseer, 2012.

[13] GAO H, TANG J, LIU H. Mobile location prediction in spatio-temporal context[C]// Nokia Mobile Data Challenge Workshop. Citeseer, 2012.

[14] BAUMANN P, KLEIMINGER W, SANTINI S. The influence of temporal and spatial features on the performance of next-place prediction algorithms[C]// Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing. ACM, 2013: 449-458.

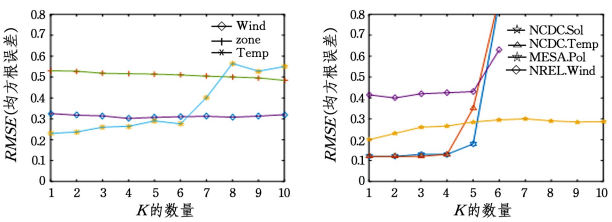
[15] TRAN L H, CATASTA M, MCDOWELL L K, et al. Next Place Prediction using Mobile Data[C]// Proceedings of the Mobile Data Challenge Workshop (MDC 2012). 2012.

[16] NOULAS A, SCELLATO S, LATHIA N, et al. Mining User Mobility Features for Next Place Prediction in Location-Based Services[C]// ICDM. Citeseer, 2012: 1038-1043.

[17] DUDA R O, HART P E, STORK D G. Pattern classification [M]. John Wiley & Sons, 1999.

（上接第 435 页）

低。即不同数据情况下 k 值与预测精度没有确定的关系，大多数情况下预测精度都随着 k 值的增加而降低。在实验时我们将近邻数量 K 的最大值设置为 10。



(a)　　(b)

图 4　K 值对预测结果的影响

结束语　本文提出 knnVAR 模型对地理时间序列进行预测，首先用寻找时空 K 近邻的方式将数据中的时间信息与空间信息融合，而后利用 VAR 模型进行多时间序列的预测。在模型中充分考虑了各个时间序列的独特性，且对近邻的数量进行了讨论，提出了抽样测试的方式来确定 k 值，以提高预测的精度。在实验中对多个实际地理传感网络中的数据进行了预测，并将预测结果与未考虑各序列独特性的 cVAR 模型的预测结果进行了对比，结果表明 knnVAR 模型比 cVAR 模型有更好的预测精度。在下一步的工作中我们将对各个序列的独特性作进一步的探讨，本文模型在寻找近邻时对整个模型的 k 值进行了确定，但每个时间序列对 k 值的选取可能会不同，需要单独讨论，这将是我们以后的工作方向。

参 考 文 献

[1] EGRIOGLU E, YOLCU U, ALADAG C H, et al. Recurrent Multiplicative Neuron Model Artificial Neural Network for Non-linear Time Series Forecasting[J]. Neural Processing Letters, 2015, 41(2): 249-258.

[2] HYNDMAN R J, KHANDAKAR Y. Automatic Time Series Forecasting: The forecast Package for R[J]. Journal of Statistical Software, 2008, 27(3): 1-22.

[3] LÜTKEPOHL H. New introduction to multiple time series analysis[M]. Springer Science & Business Media, 2005: 88-89.

[4] PRAVILOVIC S, APPICE A, MALERBA D. Integrating cluster analysis to the ARIMA model for forecasting geosensor data[C]// International Symposium on Methodologies for Intelligent Systems. Cham: Springer, 2014: 234-243.

[5] PRAVILOVIC S, BILANCIA M, APPICE A, et al. Using multiple time series analysis for geosensor data forecasting[J]. Information Sciences, 2017, 380: 31-52.

[6] BOX G E P, JENKINS G M. Time Series Analysis: Forecasting and Control[J]. Journal of Time, 2010, 31(4): 303-303.

[7] TSAY R S. Multivariate time series analysis. With R and financial applications[M]. Wiley, 2013: 1-40.

[8] KAMARIANAKIS Y, PRASTACOS P. Space-time modeling of traffic flow[J]. Computers & Geosciences, 2005, 31(2): 119-133.

[9] POKRAJAC D, OBRADOVIC Z. Improved spatial-temporal forecasting through modelling of spatial residuals in recent history[C]// Proceedings of the 2001 SIAM International Conference on Data Mining. Society for Industrial and Applied Mathematics, 2001: 1-17.

[10] SAENGSEEDAM P, KANTANANTHA N. Spatial time series forecasts based on Bayesian linear mixed models for rice yields in Thailand[C]// Proceedings of the International Multi Conference of Engineers and Computer Scientists. 2014: 1007-1012.

[11] QIN K, CHEN Y, ZHAN Y, et al. Spatial clustering considering spatio-temporal correlation[C]// International Conference on Geoinformatics. 2011: 1-4.

[12] BIRANT D, KUT. ST-DBSCAN: An algorithm for clustering spatial-temporal data[J]. Data & Knowledge Engineering, 2007, 60(1): 208-221.

[13] APPICE A, CIAMPI A, MALERBAD. Summarizing numeric spatial data streams by trend cluster discovery[J]. Data Mining and Knowledge Discovery, 2015, 29(1): 84-136.

[14] APPICE A, GUCCIONE P, MALERBA D, et al. Dealing with temporal and spatial correlations to classify outliers in geophysical data streams[J]. Information Sciences, 2014, 285(1): 162-180.

[15] REYNOLDS A P, RICHARDS G, IGLESIA B D L, et al. Clustering Rules: A Comparison of Partitioning and Hierarchical Clustering Algorithms[J]. Journal of Mathematical Modelling & Algorithms, 2006, 5(4): 475-504.

[16] ZIVOT E, WANG J. Modeling Financial Time Series with S-PLUS? [M]. New York: Springer, 2006: 296.