

【Café速递】许磊：地理学中预测的不确定性问题探索

及在美联合培养经历分享

核心提示：地理学中预测的不确定性该如何量化？在美联合培养又是一段怎样的体验？本期我们邀请到武汉大学测绘遥感国家重点实验室 2018 级博士生许磊分享他在地理学中预测的不确定性问题方面的研究成果与在美联合培养经历。

主持：马箬悦 摄影：梁天祺 摄像：李皓 文字：刘林

>>>人物名片

许磊，测绘遥感国家重点实验室 2018 级博士生，师从陈能成教授。主要研究兴趣为地理时空预测。在 *Remote Sensing of Environment* 等期刊发表论文 11 篇。曾获武汉大学学术创新一等奖、国家奖学金等荣誉奖励。

>>>报告现场

10 月 23 日晚上 7 点，测绘遥感国家重点实验室 2018 级博士生许磊做客 GeoScience Café 第 272 期报告交流活动。许磊针对地理学中预测的不确定性问题做了报告，同时也和大家分享了其在美国联合培养的相关经历，令听众受益匪浅（图 1）。



图 1 许磊作精彩报告

报告伊始，许磊以“天气预报”这一生活中常见预测问题做引，进而体现预测的重要性与普遍性，开始了本次对于地理学中预测问题的报告，主要从以下四

个方面展开分享：

第一个是地理学中的预测问题

第二个是预测不确定性及量化问题

第三个是遥感数据同化问题

第四个是在美联合培养经历

地理学中的预测问题

许磊首先从定义谈起。预测的定义是什么呢？从过去预测未来，从已知预测未知，这就是预测的基本功能。但在地理学中，很多数据都是时空数据，因此在进行预测时就成为了地理时空预测的问题。时空预测与传统的时间序列预测以及空间插值的区别就在于时空预测用到的数据是时空数据，而不是简单的时间序列，或者是简单的空间上的二维的矩阵。以降雨的预报为例，降水在空间上是连续的，因此预测是一个涉及到空间连续性的问题，其自身就是一个时空数据。



图2 报告会现场

关于数据，许磊举例说道，在今年7月份的时候，中国南方发生了非常严重的洪涝灾害，如果能够提前预测几天或者是几个小时，那么就能够为洪涝灾害做一些准备工作来减少损失。地理时空预测方法大致分为两类，其中一类是数值预报的方法，也是目前天气预报采用的方法。另外一种统计机器学习方法。

数值预报的含义就是用物理模式来模拟天气系统。物理模型的可解释性非常

强，因为它是基于物理含义的。统计和机器学习的方法，也经常用来做预测，但是是数据驱动，数据驱动的解释性就会弱一些。

统计和机器学习的这种方法分为三类，第一类就是混合时空模型，第二个是时空克里金插值，第三类就是现在比较流行的基于深度学习的时空神经网络模型。混合时空模型是在时间序列预测的基础上，在空间维度上做扩展。时空克里金插值，是在空间插值基础上在时间上做扩展。基于深度学习的时空神经网络，它本身考虑到空间上的一个维度，通过例如卷积的方式来考虑空间上的相邻关系进而预测。

预测不确定性及量化

不确定性问题分为数据与方法两类。对于预测方法来说，使用方法不一样，那么预测的结果也就不一样。

许磊把地理时空数据简单分为三类，首先是基于站点插值的数据，比如气象数据，就可以用实测站点数据来进行空间的插值，形成空间连续的面数据；第二个是基于遥感反演的数据，比如用遥感影像来观测，然后再通过一些反演模型来进行反演；第三个是再分析数据，它是基于模型和观测得到的数据，首先利用模型运行，但是可能运行结果并不准确，在有观测数据之后，可以把观测的数据同化到模型里面，这样就形成再分析的数据，所以再分析数据是基于模型和观测得到的资料。

数据具有不准确性，所以需要进行评价。基于真值的验证是较为常用的一种方法，但在一些情况下，不能获取真值，不过如果具有三种误差独立的数据，也可利用 Triple collocation 或者 Three-cornered hat 等方法进行验证。但是要注意的是这两种方法所能评价的误差是随机误差而不是系统误差。

第二个就是预测方法的不确定性，如下图所示（图 3）。图中有很多种方法，但它们的精度是不一样的。最后一行的精度是 87.6%，倒数第二行的精度是 87.1%。而在实际预测时，选择 87.6%得到的结果与选择 87.1%得到的结果可能都会正确，也可能都不正确，这就是预测方法的不确定性。

二、预测不确定性及量化

□ 预测方法不确定性

Model	Sens	Spec	PPV	NPV	F ₁	Acc (95% CI)	Diff acc (95% CI)
LDA	44.2	72.4	60.6	57.5	51.1	58.6 (55.0, 62.2)	-29.0 (-32.4, -25.6)
MNB	82.3	72.6	74.2	81.0	78.0	77.3 (73.9, 80.7)	-10.3 (-12.5, -8.1)
SVM	83.5	84.5	83.8	84.2	83.6	84.0 (80.8, 87.2)	-3.7 (-6.6, -0.7)
LSA	81.5	88.5	87.2	83.3	84.2	85.1 (83.1, 87.0)	-2.6 (-4.2, -0.9)
NN _{sum}	85.5	84.7	84.4	86.0	84.9	85.1 (81.9, 88.3)	-2.6 (-5.2, 0.1)
NN _{avg}	86.3	86.4	85.9	86.9	86.0	86.3 (84.4, 88.2)	-1.3 (-3.3, 0.6)
RF	87.0	87.1	86.6	87.5	86.8	87.1 (83.8, 90.4)	-0.5 (-2.2, 1.1)
NB-SVM	85.2	89.9	89.0	86.4	87.1	87.6 (85.2, 90.1)	*

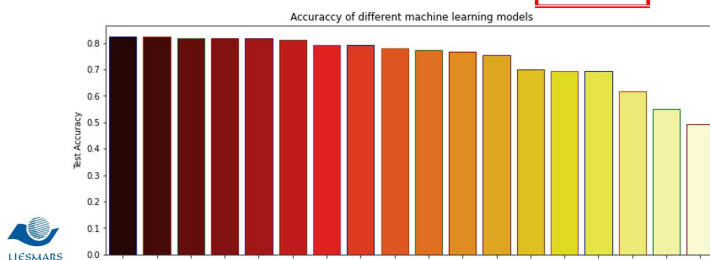


图3 预测方法的不确定性

而关于究竟选择哪一种方法，许磊提到了数值预报里提出的一种集合的思想——不特意选择一种最好的，而是生成一个集合，运用集合的方式来表达它预测的不确定性。

遥感数据同化

许磊先是介绍了数据同化的概念。用于模型模拟的地表初始条件很可能是不准确的，导致模型运行可能产生偏差。观测数据可以用来对模型模拟的状态和参数进行更新，即数据同化。之后，许磊又对较常见的集合卡尔曼滤波和粒子滤波算法等做出了深入的解释，也结合 SMAP 卫星土壤湿度数据同化的例子进行了说明。

在美联合培养经历

在报告的最后，许磊分享了自己在美国联合培养的相关情况。许磊所在的国外团队成员中有3位也来自中国，也有来自伊朗、瑞典等国的成员，平时也会说中文，在开会、作报告等正式场合还是会以英文为主，不同国家之间的文化得到了很好的交流。

申请流程方面，要在提前一年的9到12月份与外导取得联系，做提前准备；12月份开始武汉大学的预申请；次年5-6月，录取结果公布，便可以着手准备护照、签证、出国机票等事宜。

联合培养的期限可以是6个月到两年，硕士可以是一到两年。语言方面，理

论上可以不考雅思之类的语言考试，联系外导开具一个语言证明即可。

许磊提醒道，在国外的学习要做严格的学习计划，平时与学生交流更多，与导师交流机会并不是很多。美国住宿方面一般以整年进行租赁，而且退租需要联系下一租户，较为繁琐。交通方面不是很方便，尤其是校外。



图4 合影留念

>>>互动交流

提问人一：你好，看到刚刚展示的幻灯片上变量在L尺度的方差是不一样的，是不是可以计算在月份尺度、天尺度上的，尺度不同会不会对结果造成一定的影响呢？

许磊：没错，是不同的，可以通过例如一个月的尺度的方差除以总的方差从而得到在月份尺度的可能的可预测性。

提问人二：你好，如果做数据同化的话模型是不是基于一个平稳的假设、平稳的过程，在这个前提下，根据一些观测数据进行同化？如果是在突变的状况下，比如说突然爆发，效果怎么样？

许磊：个人认为，通过模型模拟，然后在后续加入观测数据是可以进行突发状况的预测的，比如干旱和洪水预警。

提问人二：做同化时，后验的采样、后验的概率预测应当如何操作呢？用什么手段实现呢？

许磊：我是利用 MCMC，由于我实验的区域比较大，所以运行速度比较慢，当然这也与计算机硬件性能有关。

提问人三：你好，在刚刚讲地理时空预测的未来时，图解没有很详细的标注，单位和数值大概分别是多少？

许磊：横轴是时间，纵轴是精度。我认为它最后肯定会达到一个极限值，极限值达不到百分之百，和具体预测的变量是有关的。举个例子，比如现在温度的预测，它可能是比较成功的，即使往后预测一个月，相对来说误差可能也比较小；但是对于降水来说可能就不太准确，两周以后就不准了。今后预测精度会不断提高，因为预测方法和计算机技术在发展，还有进步的空间。

GeoScience Café以“谈笑间成就梦想”为目标，于每周五晚 7:00 在实验室四楼休闲厅，邀请 1-4 位嘉宾，为大家带来学术报告或经验分享。报告内容包括摄影测量与遥感、地理信息系统、导航与定位服务等研究方向，听众可在报告结束后向嘉宾提问、与嘉宾交流探讨，同时每学期还会举办 2 期人文类讲座和 2 场导师信息分享会。每期报告会根据嘉宾意愿在 B 站开设直播，使不能来到现场的听众同步参与。报告 PPT 和视频会在征得嘉宾同意的情况下在 qq 群和 B 站上发布。

更多精彩内容（讲座预告、讲座回顾、报告 PPT、报告视频）敬请通过以下方式获取：



QQ群



微信公众号



B站直播