***Modeling and Predicting Learning Behavior in MOOCs***

**论文阅读报告**

电子科技大学 郭志帅

**一、论文的基本信息**

Jiezhong Qiu, Jie Tang, Tracy Xiao Liu, Jie Gong, Chenhui Zhang, Qian Zhang, and Yufei Xue. Modeling and Predicting Learning Behavior in MOOCs. In Proceedings of the Ninth ACM International Conference on Web Search and Data Mining (WSDM'16). pp. 93-102

**二、论文要解决的问题及其重要性**

利用慕课在线学习环境提供的大量的学生交互记录分析学生的学习行为，并构建一个统一的模型预测学生的学习效果。

通过收集全球的教育资源和重新构建学习环境，慕课正在引领着一场教育革命。所有学生的学习行为都发生在线上，可以完整被记录下来，这给针对学生学习行为的研究创造了前所未有的条件。但是慕课课程的低完成率，与传统教学模式不同的特征等也带来了研究上的新挑战。

以上种种，都对慕课上的用户行为分析提出了明确的需求。更重要的，如何有效的诱导用户参与课程学习和社交互动成为了一个极具意义的课题。

**三、论文提出的解决方案**

（1）采用线性回归分析检测学生统计学特征和课程选择之间的关系。其中统计学特征包括性别和教育程度。

（2）研究学习活动，包括论坛活动、观看视频和完成作业等， 并检查这些因素对用户获得结课证书可能性的影响。其中针对观看视频和完成作业定义了有效学习时间:

定义一：有效学习时间。开始和暂停终止之间的时间段为有效学习时间。如果在某一状态停留时间过长超过阈值，则自动将之后的时间归为非有效学习时间。

（3）建立潜在的动态因子图（LadFG）模型。该模型把可观测的学习行为，包括参与论坛活动、观看视频和完成作业等，结合到一个统一的框架内。

**公式化：**

定义二: 学习活动。用张量的每个元素表示在时间学生的第个活动。其中T是时间戳的个数，n是活动的个数。用表示在时间学生的所有活动。

定义三：潜在学习状态。对时间t的每一个学生，定义一个m维向量表示连续性的潜在状态，其中。张量记录了所有学生在所有时间戳的潜在状态。用表示用户在之前p个时间戳的潜在状态。

定义四：时变属性张量。 张量中代表学生。d是定义的属性个数，用向量代表学生在时间t的属性值。

对学习行为建模的目的是找到从学生属性张量到学习活动的一个映射。运用潜在学习状态作为桥梁连接两个观测变量集合。决定于，决定于。

在LadFG模型中，用。文中采用了逻辑回归模型，并将*f*定义为：

将学习活动建模为：

其中，为观测模型参数。

动态模型的输入：

潜在状态：

其中，为模型参数。

目标函数：

结合EM算法和梯度下降方法选择参数，最小化目标函数。

**四、方案效果**

利用LadFG模型进行作业等级预测和证书获得预测，结果显示该模型在相关实验中明显优于其他对比方法（逻辑回归、支持向量机、分解机）。

**五、相关研究**

（1）属性分析：关注用户的统计学特征和在线表现模式的关系。

（2）参与分析：分析用户对课程的参与度，如投入的时间等。

（3）论坛分析：针对用户在慕课平台上论坛中的行为进行分析。

**六、收获**

（1）数据挖掘首先要有大量的数据来源，本文选择针对的是慕课平台，李义萍博士的学位论文选择了微博。Android有大量的用户，下一步要学习如何从Android提取用户行为记录。

（2）类似李博士的论文，学习了从一个现实问题抽象到数学问题的建模过程。

（3）不规则不理想的数据大大减少了可以利用的数据量，例如本文中课程完成率极低，论坛活跃人数极少，以及李博士论文中微博发布频率高于要求的人数极少。而现实数据的不理想也会大大地限制模型的适用范围。