**When Choice Happens: A Systematic Examination of Mouse Movement Length for Decision Making in Web Search**

**高祎珂\_20221024\_论文笔记**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **基本信息** | 发表刊物 | SIGIR | | 发表年份 | 2021 | 第一完成单位（国内） |  |
| 作者 | Lukas Brückner，Ioannis Arapakis，Luis A. Leiva | | | | | |
| 关键词（中文） | | 鼠标光标跟踪;决策;深度学习 | | | | |
| 关键词（英文） | | Mouse Cursor Tracking; Decision Making; Deep Learning | | | | |
| **论文内容** | 解决的问题（如有实际应用场景请说明） | | 通过对鼠标光标移动建模，预测搜索者在SERPs上的选择。  其中构建了三个场景   1. 注意到一个广告 2. 关闭该页面 3. 变得沮丧   通过有效记录数据量，可以节省网络带宽和存储，提高机器学习模型的训练和部署的速度。 | | | | |
| 解决问题的方法（采用什么模型框架等） | | 递归神经网络(RNN)：   1. 使用了由Arapakis等人提供的相同的RNN模型和设置，该模型是双向长短期记忆（BiLSTM）层，有0.5dorpout(遗忘)，50个记忆单元，然后是1个全连接的神经元激活作为输出。BiLSTM层使用双曲切线作为激活函数和sigmoid函数。 2. 使用由 Brückner等人提供的相同的RNN模型和设置，该模型是2个BiLSTM层的堆栈，0.3dorpout(遗忘)，然后是1个全连接的sigmoid神经元激活层作为输出。BiLSTM层在循环步骤中使用双曲切线作为激活函数和sigmoid函数。 3. 放弃模型的任务之前没有使用过RNN，作者这里也没有使用，使用的是和abandonment model一样的结构，也即H. A. Feild, J. Allan, and R. Jones. 2010论文中作者使用的模型，将挫折预测视为一项二元任务，但是使用回归或多类机器学习进行多类预测，因为他们把挫折分为了多个等级，这篇论文中也是这样。 | | | | |
| 仍旧存在的问题（注明论文中说明的问题或自己认为存在的问题） | | 论文：设计到隐私问题，可以得到性别年龄，短鼠标序列得到少量信息不能描绘出使用者的全部轮廓。  自己：概叙列对于预测存在一定的延时性，但是论文中并没有提出。 | | | | |
| **实验内容** | 实验采用的数据集 | | 三个任务采用了三个数据集   1. the attentive cursor dataset L. A. Leiva and I. Arapakis. 2021. The Attentive Cursor Dataset. Front. Hum. Neurosci. 14 (2021). 2. L. Brückner, I. Arapakis, and L. A. Leiva. 2020. Query Abandonment Prediction with Recurrent Neural Models of Mouse Cursor Movements. In Proc. CIKM. 1969–1972.（Yahoo SERPs133 个带标签鼠标序列） 3. H. A. Feild, J. Allan, and R. Jones. 2010. Predicting Searcher Frustration. In Proc. SIGIR. 34–41.(几个SERPs上关于在浏览信息和导航查询时搜索的挫折（用户是否沮丧）的259个带标签鼠标序列)。 | | | | |
| 数据集内容是否和待解决问题模型对应 | | 不同任务使用了不同的数据集，每个数据集跟其任务都是相吻合的。 | | | | |
| 实验是否涉及实际应用场景 | | 涉及到了实际应用场景，这篇论文的实验就是基于实际应用场景展开的，对于要研究的问题，对三个实际应用场景进行建模，即是否注意到广告，是否关闭页面，是否沮丧，基于这三个场景展开实验。 | | | | |
| 实验采用的对比方法 | | 根据评价指标F1-score和AUC-ROC，通过控制变量，比较不同序列长度对模型的影响，根据实验结果做出条形图，比较图形，得出效果。 | | | | |
| 实验任务 | | 再此篇论文之前，已经有基于鼠标光标对最终SERPs选择的模型，这篇论文作者基本研究的问题是：鼠标序列长度如何影响模型性能。论文作者通过对三个场景建模进行实验，观察不同场景不同序列长度的实验结果。  在每个场景中，我们使用分层抽样对数据进行分割，使用60%的数据用于训练，10%用于验证，30%用于测试。以16个鼠标序列分批训练模型，使用Adam优化器，学习率为η= 0.0005和衰减率为β1=0.9和β2=0.999。我们设置了最大500个epoch，但通过使用10个阶段监测验证损失（二进制交叉熵）来提前停止训练；即，如果验证集上的损失在连续10个时期没有改善，则训练完成，最佳学习权重也被保存。  基于数据集的不同，三个场景使用的序列长度也不相同，对于abandonment任务，由于最大时间步只有123，所以在实验时，此场景只设计了序列时间步长为10，20，50，100，full，而没有200. | | | | |
| 实验衡量指标 | | F1-score，AUC-ROC | | | | |
| 实验说明所提出方法的优点 | | 第一次对输入RNN序列长度的研究  ,我们是第一个研究序列长度对RNN性能的影响在信息检索任务上 | | | | |
| **思考内容（阅读论文后自己思考填充）** | 论文的主要优点是什么 | | Our work is the first to provide empirical evidence for the three  paradigmatic SERP scenarios considered: predicting ads noticeability, page abandonment, and search frustration.利用前任使用的模型，测试了三个场景下鼠标序列对模型的影响，也位这三个场景提供了经验证据 | | | | |
| 论文仍然可以改进的地方是什么 | | 对于模型的选择上，本篇论文中，对于每个场景均是借用之前的人使用的模型，感觉只是做了一下整合；  对于实验的指标的选取，仅仅查看了F1-score和AUC-ROC，可以在多观察几个指标看看效果。 | | | | |
| 以此论文为出发点，如果需要你做一篇和其相关的顶会论文，你需要的资源是什么？数据，硬件，技术支持等 | | 调研另外的架构和数据集，这个做法论文作者也提了出来，可以查看其他的模型架构，本篇论文基本是基于RNN来实现的，可以去查看其他架构，看是否会得到更好的效果。  如果想要做类似研究的话，我还是需要相关领域的数据集，比如其他应用场景下的鼠标序列，这是一些带标签数据，利用这些数据和架构进行建模。  在硬件方面，本论文作者使用的是CPU @ 2.4 Hz的机器，本实验的计算量并不是过于大，所以对硬件要求不太高，足以运行模型即可。 | | | | |
| 其他想要补充说明的内容 | | 论文结论：只关注前面的一些序列（我们的视觉系统是很快的  甚至在移动鼠标之前。这尽量减少了阅读和解释页面的思考或“锻炼”，只是让眼睛做他们有效的工作，然后鼠标跟随。相反，对于预测SERP的放弃，关注最终的鼠标运动是有益的。  用户在任何时候都会感到沮丧，即使他们已经满足了他们在SERP [10]上的搜索需求，因此很难仅通过跟踪鼠标运动来预测这一结果。  综上所述，注意预测任务应关注鼠标运动序列的初始时间步长，而放弃预测任务应关注最终序列的时间步长。它只需要大约20个时间步长（大约3秒的交互数据），就可以检测到上述具有合理性能的决策过程 | | | | |