

什么是用户点击模型

【马少平 王东】

现代搜索引擎会利用各种信息来提高搜索结果的质量,其中用户点击模型在结果排序中具有重要作用。那么什么是用户点击模型呢?

当一个用户输入查询词之后,会根据自己的需要对搜索引擎给出的结果进行选择性的点击,用户之所以会点击某个搜索结果,一定是该用户认为这个搜索结果可能对自己是有帮助的。所以,点击可以认为是用户对搜索结果的一次投票,当多数用户点击相同的搜索结果时,说明这个搜索结果是比较重要的,下次搜索时可以提高这一结果的排名,并在网页靠前的位置显示。

那么,是不是被点击越多的结果就越好呢?也未必。用户对结果的点击受多种因素影响,除了是否相关外,一个重要的前提是用户看到了这个结果,才有可能点击。排在前面的结果被用户看到的可能性大,排在后边的结果被用户看到的可能性会小一些,因此前面的结果天然会被点击的多一些。换句话说,一个点击量少但排在后面的结果不一定就比点击量多但排在前面的结果差,这就是所谓的位置偏置。除了位置偏置以外还有很多其他的偏置,例如,人们在浏览搜索结果时,对网页不同位置的关注度是不同的。图1给出了利用眼动仪得到的用户关注位置的热力图,不同的颜色反应了用户对搜索结果页面不同区域的关注程度。可以看到,用户对不同位置的关注度不同,这种关注习惯上的差异显然也会影响用户点击的结果。



图 1: 眼动仪收集的用户查看热力图

如何在具有各种偏置的前提下,依然可以基于用户的点击行为来判断搜索结果的相关性呢?

这就是用户点击模型所做的工作[1]。所谓用户点击模型,是用来描述用户从开始搜索到搜索结束过程中点击行为的发生过程的模型,他利用大量用户的历史点击信息对用户的兴趣和行为进行建模,消除可能存在的偏置。基于这一模型,可以判断网页上每条结果的“真实”价值,这些信息反馈给搜索引擎,可以对网页进行更有效的排序。

值得说明的是,用户点击行为非常复杂,因此在建立点击模型时需要做出各种假设。不同的假设可以得到不同的点击模型,有的简单有的复杂,其效果也各有不同。常用的用户点击模型包括级联模型、依赖点击模型、用户浏览模型、动态贝叶斯模型、点击链模型等。不论哪种点击模型,其根本目的都是透过“点击”这个基本操作来判断用户真正感兴趣的东西,从而改进搜索引擎的输出质量。

[1] 王超,刘奕群,马少平. 搜索引擎点击模型综述[J]. 智能系统学报, 2016, 11(6): 711-718.