对毕设课题的理解

11612126 李可明

问题:

Cover Ratio Maximization (user distribution ):

Given a dataset , product group , product creation budget , a positive integer and a creation cost function , introduce a new product such that and is maximized

对问题的理解:

我对D的理解是product 的一个集合， 是离散的

P 是另一个product 的集合， 也是离散的

某些product in P 在P respect to some user preference vector能排前k， 覆盖了一部分user preference vector

我们要找一个新的满足约束C(p)<B 的 product p， 使得P{p} 覆盖最多的user preference vector

解决问题路线:

可以结合k-hit query 的思路对cell tree 修改， 这里应该可以有一个baseline, 原先希望这周能想出baseline的思路但还有些问题。

因为k-hit query 是top1， 用到一个k-hit query 自己定义凸包（convex hull）， 只有凸包上的点才有可能会排top 1。

我们的要做的是topk， 这意味着我们

1. 首先要找到哪些点有可能会排top k(或者说哪些点根本排不了tok), 找到这个解集

这里可能会修改k-hit query的某些部分， 由于是topk 不是top1 还会用到修改的cell tree

1. 根据约束C(p)<B 挑选合适的p

这里一般来说就是在约束边界上找解， 这意味着第1步根据这个约束能有一些剪枝

1. 由于P已经覆盖了一部分user preference space, 找到cover ratio maximization的新的p 并不简单， 但一般来说解就在约束条件上

但是在第2步中由于找到的解集肯定是连续的区域解集， 这意味着我们要在连续的product space 找到一个解， 问题已经不是单纯的离散的product space

Baseline 暂时还有点问题， 需要点时间给出来

计划是希望

1. 寒假能将baseline的思路， 每一步的证明完善以及代码基本框架做出来
2. 回来后跟老师讨论baseline 的理论正确性， 哪些地方要改
3. 回来第一个月(2月)根据跟老师讨论的结论得到正确的baseline或者更好的算法
4. 回来第二、三个月(3月， 4月)是代码实现跟测试
5. 4月、5月跑实现， 写论文

问题2

Given a dataset , a user preference set , a cover percentage , a positive integer and a creation cost function introduces a new product such that the cover ratio

and the creation cost is minimized

对问题的理解：

D是product space的数据集， 是离散的

W 是服从某种分布的user preference vector的集合

Product space(这个是连续的空间) 上找到所有满足覆盖率不低于 的p， 在这些p中筛选代价creation cost 最小的一个解