**MAB for Task Allocation**

1. 网络的结构是变化的，如何确定K值？

能否找到应用场景是一个节点周围有固定的K个邻居，节点的周围K可以是不一样的，但是每个节点的周围个数是不变的。

记录在一段时间内某个节点遇到哪些节点作为这个节点的邻居（K个可以联系的邻居，电话或者短信）。假设该节点有T个task要发布，每一轮发布一个task，这个时候该节点从邻居节点中选择较好的节点完成，使自己的reward最大化。直接考虑单跳，k个邻居看作是k个arm。

1. 对于reward如何定义？

任务发布者从邻居里面选择节点来执行任务(发短信，电话等方式)，如果该任务被正确合格的完成了，那么reward=1。每个worker 的完成工作的能力不同，每个worker 的工作负载不同，这些都是未知的。

或者把reward转换成regret来度量。通过最小化regret来实现目标。

1. 转化成regret，如何去确定算法的复杂度？证明是对数解，如何证明
2. 每个arm的reward分布?
3. 算法的选择：ε-greedy、Softmax、UCB、Thompson Sampling?

选择Thompson Sampling的方式，先了解下Thompson Sampling的做法，然后改进使其适用于本问题。

1. 任务的发布针对蜂窝网络(Celluar network)。
2. 参考于：

A Multiarmed Bandit Incentive Mechanism for Crowdsourcing Demand Response in Smart Grids

通过设计一些激励机制，激励worker来接受任务，然后相当于任务发布者有个utility=R-payment

1. 目标是如何从众多的承包商中选择较优的承包商。使任务发布者的Regret最小。

作图

1. 每轮间隔可以设置为1小时或者30分钟，在某一轮中选择一个arm，如果选择了，那么接下来就要分配payment，如果没有成功分配payment，那么在下一轮就接着选择arm。如果选择失败，在下一轮接着执行TS算法选择task。画图表示不同的round间隔下的不同情况。
2. Task定义个delay (T)
3. Regret performance，理论分析证明regret的bound，最后实验有个regret，画图表明两个之间的逼近。
4. Task的总数对累计regret的影响，或者task产生的速率对累计regret的影响。
5. Context的取值：1). Worker 的bid取值； 2) worker的buffer size大小； 3) 某类任务的完成能力； 4) 某类别完成的次数； 5) energy; 6) centrality