# Problem identification checklist

|  |  |
| --- | --- |
| 1. Title  (no more than 10 words) | - |
| 2. Scope  (concrete, specific, narrow down) | Efficiently recommend piecework to Requester of Crowdsourcing online market, ZHUBAJIE, [-based on system reputation and behavior of workers ] |
| 3. Problem statement | 1. Relationship: quality of piecework and user behavior 2. Relationship: requester review distribution 3. 任务、稿件对合格率的影响关系 |
| 4. Definitions  (write out mathematically) | TO BE DONE  Change pairwise (or listwise) model of learning to rank.   1. Define a method to modify weight for one choice of a task, which may not create distinguish advantage for the picked one to others. 2. May definite an evaluation method(or) for this situation.   3. Definite the situation to make it easier to describe . |
| 5. Meaning/Importance/ Advantage/Applications  (maybe concrete or broad, like social tools, high accuracy, low energy, high coverage, enriched capability, more scalable, privacy preservation, user experience, low cost, low infrastructure requirement, non-intrusive; the meaning may be enabling apps) | Effectively Recommend piecework to requester to save their time.  [Impove Accuacy]  <!--  behavior 相关不在此次工作  Advantage?  As the model based on user behavior and reputation, not based on the content,  so the application could be built beyond their system.  The application may be a crx file on chrome  -- > |
| 6. List of entities  (help you identify possible problems to explore) | 1. 任务HIT (human intelligent task) 2. 稿件 piecework |
| 7. Related work  (if any; limitations) | [WSDM’15] Using Employer Implicit Judgements To Infer Worker Reputation  limitation.  In the Elo. rating system on which their pairwise ranking bases, A beat B show A is real better then B. But in our task, A beat B is because of that the employer prefers A, or he like both but made a tough choice randomly.  On the whole zhubajie platform , only employers have the right to make the choice, so we may not treat one choice as an error.  For example, a porn will be vote top video when voting a best video.  (年前，腾讯新闻就把一个黄色内容的新闻稿，机器自动筛选成为了新闻，发布出来，就是因为voting高) |
| 8. Challenges  (system requirements; may not be necessary if the proposal is novel and exploring; you can instead mention limitations of related work) | 1. small false negative [definite an evaluation method] 2. ranking method/ evaluation 3. how to find ground truth and how to reflect the relation between piecework |
| 9. Any simple but effective solutions | 1. 在单用户多次submit的场景下，可以通过添加黑名单来完成.   (需要考虑提交任务的用户数量) [计件类别] |
| 10. Ideas/Intuitions  (you can write down any intuition as a detailed algorithm [not general description] - if, else, a <- a+b) | To predict a group of piecework which may have much more probability to be chose, instead of predict a rank list. |
| 11. Contributions/Novelty | 1 method/model meaning  2 similar to intuitions[10] |
| 12. System flowchart  (block diagram; specify clearly the input and output of each block component) | 输入: 任务以及所有交稿稿件。  任务的金额、每个人的最高件数(可选)，发布时间，类目  交稿信息: 交稿时间、交稿人id,交稿时间，交稿日, 服务商系统特征(金牌),状态  输出目标:  每个稿件的被接收概率排名List  或者 某几个稿件被接收的概率>>其他稿件  //稿件排序，排序目标:让发布者审阅接受率更高。 |
| 13. Outline  (section titles) |  |
| 14. Data collection  (how to? how long? enough? miniature? convincing? data collection also for contrasting methods?) [If not easy or cannot (e.g., mobile crowdsourcing), do not proceed] | 1. 抓取 zhubajie 网站数据   任务数据;  交稿数据  单一类别   1. 属性描述 用什么写什么 2. task任务表 任务的金额、每个人的最高件数(可选)，发布时间，收稿周期 3. piecework 交稿 |
| 15. Evaluation design  (at least 3 aspects; the baseline methods compared - may be yours as well; performance metrics [must be reasonable]; draw a table to help you determine how to set parameter values to be examined) | Baseline   1. rank by time 2. rank by system reputation 3. pairwise rank based on Elo ranking system |
| 16. Figures  (flowchart; data examination; data correlation; experimental results) | 统计数据   1. piecework-time distribution for different qualities |
| 17. Code design  (block diagrams; break code into sections; based on your evaluation goals and figures to plot, set intermediate variables to record the desired output; otherwise, you may need to rerun the same experiment multiple times) |  |
| 18. Analysis and discussion  (What are the benefits of your approach compared with existing or intuitive ones? What are the possible extensions? If respective conditional probabilities are simple, derive the marginal to highlight your model's complexity, difference, or novelty.) |  |