

**Personalized and Diverse Task Composition in Crowdsourcing**

**Abstract:**

We study task composition in crowdsourcing and the effect of personalization and diversity on performance. A central process in crowdsourcing is task assignment, the mechanism through which workers find tasks. On popular platforms such as Amazon Mechanical Turk, task assignment is facilitated by the ability to sort tasks by dimensions such as creation date or reward amount. Task composition improves task assignment by producing for each worker, a personalized summary of tasks, referred to as a Composite Task (CT). We propose different ways of producing CTs and formulate an optimization problem that finds for a worker, the most relevant and diverse CTs. We show empirically that workers’ experience is greatly improved due to personalization that enforces an adequation of CTs with workers’ skills and preferences. We also study and for malize various ways of diversifying tasks in each CT. Task diversity is grounded in organization studies that have shown its impact on worker motivation [33]. Our experiments show that diverse CTs contribute to improving outcome quality. More specifically, we show that while task throughput and worker retention are best with ranked lists, crowdwork quality reaches its best with CTs diversified by requesters, thereby confirming that workers look to expose their “good” work to many requesters.

**Existing System:**

Existing literature and a thorough examination of crowdsourcing forums such as TurkerNation,1 reveal that workers spend non-negligible time discussing how to best select tasks depending on one’s goals, which requesters to ban, and which qualification levels are required for the latest tasks on AMT. This calls for re-thinking task assignment and developing an approach to find the tasks that best fit workers’ preferences.

Recent research on task assignment can be characterized as either requester-centric or platform-centric, whereby tasks are proposed to workers in order to maximize task throughput, contribution quality, and cost. To improve the workers’ experience, we propose to generate summaries of relevant tasks using CTs. We study how different optimization choices when building CTs affect workers’ performance.

**Proposed System:**

We introduce the problem of producing personalized and diversified summaries of tasks for a given worker. Our problem formulates building a set of K valid CTs that maximize representativeness, diversity and personalization.

We map our objective function into a fuzzy clustering problem solved by the fuzzy c-Means algorithm to seamlessly integrate our optimization goals.

We run thorough user studies and online deployments with real workers on AMT and explore the impact of CTs on task throughput, worker retention, and crowdwork quality.