Please contact the authors to obtain the full PDF of this paper. The abstract and contact of the authors is below.

Title

A Unified Framework for Human-Powered Categorization

Abstract

We consider the problem of utilizing human power to categorize a large number of objects. In this problem, we are given an initial category hierarchy and a set of objects, and our goal is to find the most efficient strategy that asks the crowd questions in order to assign each object to its appropriate category in the hierarchy, such that the long-term expected monetary cost is minimized. We develop a *machine-crowd* hybrid framework and propose an *online* algorithm, in which we can gradually learn the category distribution and adaptively decide an effective order of asking questions. We prove that even if the true category distribution is known in advance, the problem is computationally intractable. We develop a natural approximation algorithm, and prove that it achieves an approximation factor of 2. We also show that there is a fully polynomial time approximation scheme for the problem. Furthermore, we adopt the Follow the Perturbed Leader strategy to guarantee that the framework achieves nearly the same performance as the offline optimal strategy. We evaluate the effectiveness and efficiency of our algorithms on the real-world crowdsourcing platform.

Authors

Xian Wu (wuxian12@mails.tsinghua.edu.cn)
Jian Li (lijian83@tsinghua.edu.cn)
Guoliang Li (liguoliang@tsinghua.edu.cn)