Crowdsourced Semantic Matching of Multi-Label Annotations

Problem Definition

Taxonomy $S$

- 喜, 好, 安, 怒, 哀, 怖, 恥, 厌, 昂, 惧

Mapping from set to set

Taxonomy $T$

- Joy, Fear, Anger, Surprise, Disgust, Sadness

Optimal Mapping $f$: $f(s) = \text{argmax}_{t} \Pr(t|s)$ (s $\subseteq S$, t $\subseteq T$)

Proposed Approach

Instances with labels of Taxonomy $S$

- Instance 1, $s_1$
- Instance 2, $s_2$
- Instance $i$, $s_i$

Learn mapping $f$ from obtained examples

Crowdsourcing Annotators

Annotate with Taxonomy $T$

Statistical Methods

1. Conventional method

2. Proposed cascaded method