

Оценка «добросовестности» экспертов при проведении массовых олимпиад

О. В. Малюгина

Московский физико-технический институт (государственный университет)

Институт проблем передачи информации им. А. А. Харкевича РАН

При проведении многих конкурсов и олимпиад, в которых задания оцениваются людьми (далее экспертами), стоит задача отсеять «недобросовестных» экспертов. Некоторые эксперты сознательно занижают оценки всем работам, кроме работ своих знакомых. У многих экспертов нет желания внимательно проверять работы, и они всем ставят высокие (или наоборот низкие) оценки.

В данной работе решалась задача выявления подобных экспертов для конкурса «Ярмарка идей». Основными сложностями данной задачи являются:

- небольшое количество экспертов, проверивших каждую работу (3-6);
- небольшое количество работ, проверенных каждым экспертом (7-20).

Каждая работа оценивалась по 8 критериям, то есть за работу ставилось 8 оценок.

Для решения данной задачи применялись два метода.

1. Рассматривалось, насколько сильно оценка данного эксперта каждой проверенной им работы отличается от оценок других экспертов этой работы. Учитывалось существование спорных критериев у работы – критериев работы, оценки экспертов которых очень сильно различались. Данные критерии вносили меньший вклад в оценку эксперта.

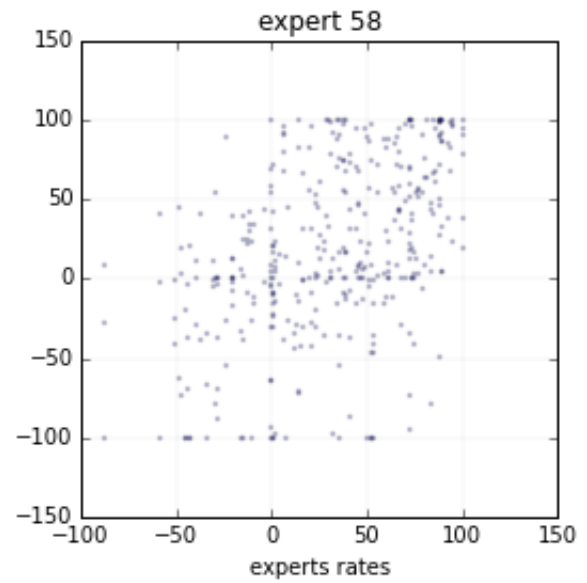
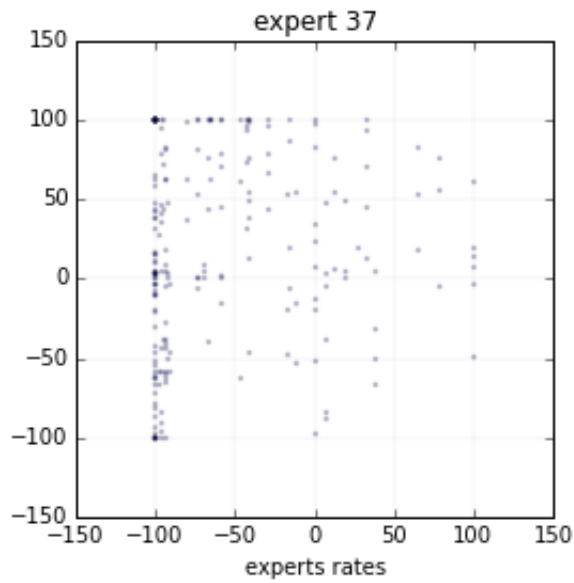
Для каждого i -го эксперта подсчитывалась следующая характеристика:

$$\sum_{w,c \in w_i} (rate_w^{ic} - \frac{1}{|j \in j_w|} \sum_{\substack{j \in j_w \\ j \neq i}} rate_w^{jc}) * s_{wc}, \text{ где}$$

- $rate_w^{ic}$ – оценка критерия c работы w экспертом i ;
- $|j \in j_w|$ - количество экспертов, проверивших работу w ,
- s_{wc} – значимость критерия c работы w (насколько он спорный).

Далее рассматривалась распределение данной характеристики для всех экспертов и эксперты, чьи характеристики были больше некоторого порога, признавались «недобросовестными».

2. Рассматривалась корреляция между оценками данного эксперта и всеми оценками других экспертов поставленных за работы, проверенные этим экспертом. Ниже представлены распределения оценок «недобросовестного» эксперта 37 и эксперта 58 (по оси x отложены оценки данного эксперта за некоторые работы, по оси y оценки других экспертов, поставленные за те же работы). Далее подсчитывалось значение корреляция между ними, и на его основании делался вывод о недобросовестности эксперта.



Данные методики обнаружили два пересекающихся множества «недобросовестных» экспертов. Оценки экспертов из их пересечения были удалены, и оценка работ проходила без них. Также были заранее известны некоторые «недобросовестные» эксперты, и большинство из них оказалось в данном перечении.

Литература

1. Вьюгин В.В. Математические основы теории машинного обучения и прогнозирования. М.: 2013. - 387 с.