

МОДЕЛЬ И МЕТОД ОЦЕНКИ КАЧЕСТВА ВКЛАДА УЧАСТНИКА СИСТЕМЫ СОВМЕСТНОЙ РАЗМЕТКИ ИЗОБРАЖЕНИЙ

А. В. Пономарев^а, канд. техн. наук, старший научный сотрудник, ponomarev@iias.spb.su

^аСанкт-Петербургский институт информатики и автоматизации РАН, 14-я линия В. О., 39, Санкт-Петербург, 199178, РФ

Постановка проблемы: масштабные человеко-машинные системы, посредством сети Интернет вовлекающие в процесс обработки информации сообщества людей, неоднородные по навыкам и уровню мотивации участников, в последнее время находят применение в широком классе приложений. В таких системах актуальной является проблема оценки ожидаемого качества вклада участника (такая оценка может использоваться, например, для отсеивания некомпетентных или небрежных участников и поощрения наиболее аккуратных). **Цель:** разработка метода оценки ожидаемого качества вклада участника системы совместной разметки, использующего только ненадежные и неполные данные, получаемые от других участников системы (без знания истинных результатов). **Результаты:** предложена математическая модель системы совместной разметки изображений (включающая модель участника), а также метод оценки ожидаемого качества вклада участника такой системы. Предложенный метод основывается на сопоставлении результатов разметки различных участников на одинаковых изображениях и является модификацией метода попарных сравнений, в которой вместо отношения предпочтения используется характеристика доминирования, определяемая специальным образом (показывающая, какой из участников дает более полные ответы). Оценки ожидаемого качества вклада участников предложено получать как положительный собственный вектор матрицы со значениями попарных характеристик доминирования. В ходе имитационного моделирования процесса совместной разметки показано, что предлагаемый метод действительно позволяет оценить качество вклада участника без использования эталонных результатов (при условии соответствия поведения участника предлагаемой модели). **Практическая значимость:** полученные результаты могут быть использованы при разработке систем, основанных на использовании скоординированных усилий участников сообщества (и, в первую очередь, систем совместной разметки).

Ключевые слова — краудсорсинг, крауд-вычисления, разметка изображений, обработка изображений, коллективный интеллект.

Цитирование: Пономарев А. В. Модель и метод оценки качества вклада участника системы совместной разметки изображений // Информационно-управляющие системы. 2018. № 4. С. 45–51. doi:10.31799/1684-8853-2018-4-45-51

Citation: Ponomarev A. V. Model and Method for Contributor's Quality Assessment in Community Image Tagging Systems. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2018, no. 4, pp. 45–51 (In Russian). doi:10.31799/1684-8853-2018-4-45-51

Введение

Технология краудсорсинга (или крауд-вычислений) в настоящее время широко используется для решения самых разных задач (см., например, [1–5]). В частности, одним из достаточно распространенных применений этой технологии является обработка изображений — интерпретация содержимого изображения, снабжение его описанием и т. п. Это требуется как при подготовке эталонных наборов данных для машинного обучения [6, 7], так и непосредственно в информационно-поисковых системах, поскольку позволяет дополнить ресурсоемкие и сложные алгоритмы обработки изображений алгоритмами текстовой обработки или алгоритмами обработки (частично) структурированной информации [8–10].

Вместе с тем привлечение человека к задачам обработки информации в рамках крауд-вычислений связано с набором принципиальных ограничений [11]: низкой скоростью обработки

информации, необходимостью мотивации, возможностью ошибок или даже целенаправленного искажения информации.

В статье предлагается математическая модель поведения участника системы совместной разметки объектов, т. е. системы, участники которой на основе анализа объектов снабжают их метками («тегами»), принадлежащими некоторому заранее известному множеству. Предлагается также метод оценки ожидаемого качества вклада участника (под вкладом в данном случае понимаются результаты разметки, полученные от участника). Предлагаемые в данной статье метод и модель демонстрируются на примере Bibtaggers — системы коллективной разметки фотографий массовых спортивных мероприятий на базе немотарного стимулирования, описанной в работе [12]. Система реализована в виде веб-сервиса, куда загружаются ссылки на альбомы с фотографиями мероприятия, опубликованными на каких-либо внешних сервисах размещения фотографий (Яндекс.Фотки, ВКонтакте). Участники меропри-

ятия получают возможность осуществлять поиск своих фотографий по номеру после того, как «разметят» несколько предложенных системой изображений (т. е. укажут, какие на них присутствуют номера участников). Мотивация к использованию такой системы связана в первую очередь с облегчением поиска фотографий. Существуют также попытки применить технологию компьютерного зрения и OCR для решения этой задачи [13, 14], однако в силу того, что нагрудные номера являются гибкими (следовательно, могут быть деформированы) и на фотографии могут быть частично закрыты, полностью автоматические решения зачастую не обеспечивают желаемого качества распознавания. Применение же системы совместной разметки Bibtaggers показало, что, как правило, к разметке удается привлечь достаточно участников, а качество получаемой разметки (измеряемое посредством точности, полноты и F-меры) оказывается выше, чем при машинной разметке (см. сопоставление [12, 15] и [13]).

Несмотря на то, что в целом качество разметки с помощью краудсорсинга оказывается удовлетворительным (и выше, чем в известных полностью автоматических решениях), качество результатов, получаемых от отдельных участников, конечно, значительно варьируется. Это связано с тем, что естественным стимулом участника в ходе разметки является минимизация прикладываемых усилий, что неизбежно влияет на качество результата. Поэтому актуальной является задача оценки ожидаемого качества результата, получаемого от участника. Эта информация в дальнейшем может использоваться как для блокировки участников, осуществляющих низкокачественную разметку, так и для поощрения добросовестных участников (или обращения к ним для разметки сложных изображений, разрешения противоречий). Задача оценки ожидаемого качества вклада участника осложняется тем, что истинные ответы в ходе разметки неизвестны, поэтому оценку необходимо проводить на основе, в общем случае, неполных и недостоверных данных, получаемых от участников.

Краткий обзор методов обеспечения качества результатов в системах крауд-вычислений

В связи с уже упомянутыми выше особенностями крауд-вычислений проблема обеспечения качества результатов является одной из наиболее острых, привлекающей внимание исследователей в этой области.

В статье [16] на основе проведенного обзора литературы выделено шесть ключевых направ-

лений, по которым могут быть сгруппированы методы обеспечения качества:

1) методы согласования (consensus) получили широкое распространение в тех вариантах организации краудсорсинга, где *недоступна априорная информация об исполнителях* и о заданиях. Общий принцип их работы заключается в том, что каждое задание выполняется несколькими участниками системы, затем полученные результаты обобщаются (согласуются) для получения итогового результата [17–20];

2) методы проектирования потока работ можно рассматривать как своего рода развитие методов согласования, в частности, позволяющее рассматривать системы с разнородными заданиями и динамически формируемым потоком работ [21, 22];

3) методы централизованного назначения работ предполагают, что для каждого участника известна ожидаемая эффективность (качество результатов, время выполнения), и строят распределение заданий таким образом, чтобы максимизировать общую эффективность выполнения набора заданий [23];

4) теоретико-игровые методы позволяют оценить ту или иную схему функционирования системы крауд-вычислений исходя из представления о поведении участника как рационального агента, стремящегося к максимизации собственной пользы [24];

5) методы, основанные на учете свойств заданий, предлагают использовать свойства заданий или их компьютерной обработки для повышения итогового качества результатов крауд-вычислений [25];

6) методы, основанные на анализе действий пользователя и воздействии на него, используют мониторинг действий пользователя и разнообразные психофизические индикаторы для выявления недобросовестного поведения [26].

Предлагаемый в данной статье метод относится к методам согласования в том смысле, что он не использует дополнительную информацию, кроме меток, получаемых от различных участников, что обеспечивает универсальность данного метода. Вместе с тем особенностью сервиса Bibtaggers (и многих других систем, основанных на немонетарном стимулировании), препятствующей применению существующих методов согласования, является необходимость значительного дублирования выполняемых заданий. То есть чтобы распространенные алгоритмы согласования (DS [19], GLAD [20] и др.) давали надежные оценки качества, необходимо, чтобы каждое задание было выполнено как минимум пятью участниками (обычно рекомендуется более десяти). Это оказывается возможным при разметке с помощью платформ краудсорсинга, поддерживающих монетарное стимулирование (Amazon Mechanical Turk, Яндекс.

Толока), потому что финансовое вознаграждение привлекает большое количество участников, которые могут посвящать разметке значительное время. В то же время практические эксперименты с разметкой на базе немонетарного стимулирования (и, в частности, с веб-сервисом Bibtaggers) показали [15], что с учетом типичной динамики заинтересованности участников каждое задание может быть выполнено только 2–3 раза.

Модель процесса совместной разметки изображений

Для оценки метода определения ожидаемого качества вклада участника была предложена формальная модель процесса совместной разметки изображений. Параметры модели были оценены в результате анализа данных, полученных при разметке изображений с помощью системы Bibtaggers.

Параметры модели. Пусть набор данных для разметки состоит из N изображений. На изображениях могут присутствовать целевые объекты (в случае системы Bibtaggers — нагрудные номера спортсменов). Задача каждого участника разметки заключается в том, чтобы снабдить изображение метками, соответствующими присутствующим на нем целевым объектам. Пусть b_i — множество целевых объектов на изображении i , а n_i — их количество ($n_i = |b_i|$).

Идентификация целевого объекта может требовать различных усилий, в то время как участники разметки обладают различным стремлением прикладывать усилия. Опишем это формально следующим образом. Пусть участник j характеризуется *аккуратностью* $a_j \in [0, 1]$. Чем выше значение a_j , тем большие усилия склонен прилагать участник и тем больше целевых объектов он способен обнаружить на изображении. Пусть также $e_{i,k}^{req} \in [0, 1]$ — уровень *аккуратности*, необходимый для идентификации объекта k на изображении i . Таким образом, в ходе исследования изображения i участник j получает множество меток

$$l_{i,j} = \{k \mid k \in b_i \wedge a_j \geq e_{i,k}^{req}\}.$$

Параметры b_i , n_i , a_j и $e_{i,k}^{req}$ неизвестны. Единственные данные, оказывающиеся доступными, — это множества меток $l_{i,j}$. На основе анализа множества меток можно пытаться получить оценки неизвестных параметров (это, в частности, делают алгоритмы согласования, основанные на EM-алгоритме), однако, как уже отмечалось, для этого необходимо иметь значительную избыточность при обработке участниками изображений. Целью данной работы является создание метода анализа меток $l_{i,j}$, позволяюще-

го получать оценки h_i , которые были бы сильно связаны (в смысле корреляции, взаимной информации или еще как-то) с аккуратностью a_j . Тогда на основе сопоставления оценок h_i можно будет принимать решения о том, как соотносятся усилия, прикладываемые различными участниками.

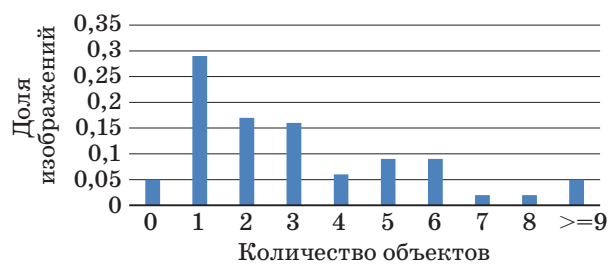
Генерация наборов данных для экспериментов. Поскольку проведение непосредственных экспериментов с системой совместной разметки оказывается затруднено, экспериментальное исследование предлагаемого метода оценки качества вклада участников производилось с помощью имитационной модели процесса разметки, на вход которой подавались сгенерированные наборы данных о размечаемых объектах.

При генерации наборов данных были использованы описанные выше параметры. Значения параметров выбирались на основе анализа истории работы сервиса (в том числе полностью размеченного вручную набора данных, описанного в работе [12]).

Количество изображений с одного мероприятия (N), размечаемых с помощью сервиса, обычно варьируется в диапазоне 1000 ÷ 7000.

Количество целевых объектов на одном изображении n_i в значительной степени зависит от стиля работы фотографа; на основе анализа исторических данных сформировано эмпирическое распределение (рис. 1), которое и было использовано при генерации наборов.

Параметры a_j и $e_{i,k}^{req}$ оказываются связанными достаточно тесным образом: одни и те же результаты можно получить, подвергая их одинаковым преобразованиям с помощью любой монотонной функции. Поэтому было принято решение при генерации наборов для выбора значения $e_{i,k}^{req}$ использовать непрерывное равномерное распределение $U(0, 1)$, а для выбора значения a_j — треугольное распределение с модой a^m , найденной таким образом, чтобы итоговое качество разметки, полученной при данном наборе значений параметров (F-мера множества полученных меток



■ **Рис. 1.** Распределение количества целевых объектов на изображении

■ **Fig. 1.** Distribution of the number of target objects on an image

по отношению к множеству эталонных), было сопоставимо с качеством, получаемым на практике (см. [12, 15]).

Метод оценки ожидаемого качества результатов, получаемых от участников

Предлагаемый метод основывается на сопоставлении результатов различных участников на одинаковых изображениях и является модификацией метода попарных сравнений, в которой вместо отношения предпочтения используется характеристика доминирования, определяемая специальным образом, а именно: эта характеристика призвана показывать, какой из участников предоставляет более полные ответы.

Для пары участников u и v характеристика доминирования является числом $d_{uv} \in (0, \infty)$, которое принимает значение больше 1, если есть основания считать, что участник u дает более полные ответы, и меньше 1, если есть основания считать, что участник v дает более полные ответы, причем $d_{uv} d_{vu} = 1$.

Эксперименты проводились с двумя способами вычисления характеристики доминирования.

Первый способ (d_{uv}^1). Пусть t_{uv} — множество изображений, размеченных и участником u , и участником v . Будем говорить, что разметка изображения $t \in t_{uv}$ участником u доминирует над разметкой этого изображения участником v тогда и только тогда, когда множество $l_{v,t}$ является строгим подмножеством $l_{u,t}$:

$$u \succ^t v \Leftrightarrow l_{u,t} \supset l_{v,t}.$$

Пусть $t_{u>v} = \{t \mid t \in t_{uv}, u \succ^t v\}$ — множество тех объектов t , разметка которых участником u доминирует над их разметкой участником v . Очевидно:

$$t_{u>v} \cap t_{v>u} = \emptyset; \quad t_{u>v} \cup t_{v>u} \subseteq t_{uv}.$$

Тогда

$$d_{uv}^1 = \frac{1 + |t_{u>v}|}{1 + |t_{v>u}|}.$$

Второй способ (d_{uv}^2). Этот способ тоже основывается на введенном выше понятии доминирования разметки, но характеристика доминирования рассчитывается по следующей формуле:

$$d_{uv}^2 = 1 + \frac{|t_{u>v}| - |t_{v>u}|}{|t_{uv}| + 1}.$$

Оценки аккуратности участников предлагается получать как положительный собственный век-

тор матрицы со значениями попарных характеристик доминирования (d_{uv}^1 или d_{uv}^2). Легко убедиться, что матрицы характеристик доминирования, вычисленные любым из рассматриваемых способов, являются квадратными и положительными. Следовательно, согласно теореме Фробениуса — Перрона [27], у них есть собственные векторы со строго положительными координатами.

Экспериментальное исследование метода

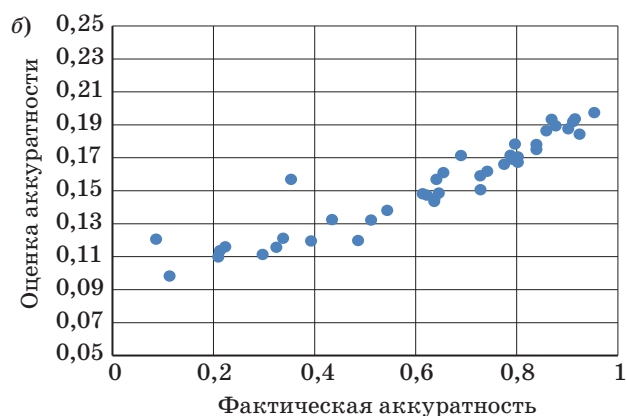
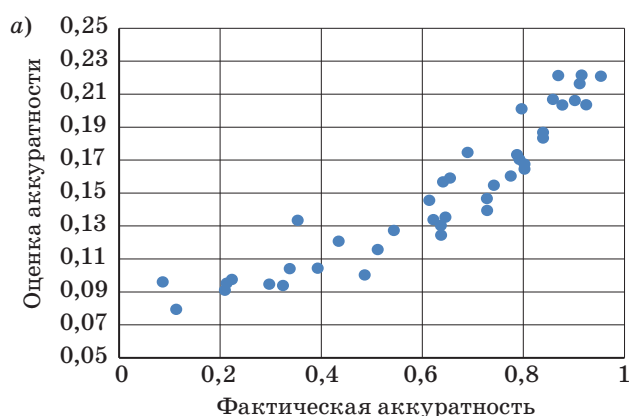
Применим предложенный метод для оценки аккуратности участников на наборе данных, сгенерированном при помощи модели, описанной выше. Значения параметров, использованные при генерации, сведены в таблицу.

Помимо этих параметров, для полного определения процедуры необходимо задать желаемый уровень избыточности — сколько участников должны обработать каждое изображение. В данном эксперименте уровень избыточности был установлен равным двум, т. е. каждое изображение обрабатывалось ровно двумя участниками, причем каждый участник обрабатывал только 50 изображений.

Связь между оценками аккуратности участников, полученными предложенным методом (с различными способами вычисления характеристики доминирования) и фактическими значениями аккуратности (задаваемыми при генерации набора данных) показана на рис. 2, а и б. Видно, что при обоих способах вычисления характеристики доминирования между оценкой аккуратности и фактической аккуратностью наблюдается сильная зависимость. Для численной оценки степени зависимости воспользуемся коэффициентом ранговой корреляции Спирмена, обладающего свойством инвариантности по отношению к монотонным преобразованиям, т. е. показывающим, насколько хорошо зависимость между двумя величинами может быть описана монотонной функцией (а поиск монотонной функции оценки аккуратности и является задачей, решаемой в данной статье). Коэффициенты Спирмена для предлага-

- Значения параметров, использованные при генерации набора данных
- Parameter values used to generate the dataset

Параметр	Значение
N	1000
n_i	В соответствии с распределением (см. рис. 1)
$e_{i,k}^{req}$	$U(0, 1)$
a_j	В соответствии с треугольным распределением (0; 0,82; 1)



■ **Рис. 2.** Связь между фактической аккуратностью и оценкой аккуратности участника, вычисленной с помощью характеристик доминирования d_{uv}^1 (а) и d_{uv}^2 (б)

■ **Fig. 2.** Relationship between the actual accuracy and the econtributor's accuracy estimated with a help of domination characteristics d_{uv}^1 (а) and d_{uv}^2 (б)

емого метода с различными способами вычисления характеристик доминирования приближенно равны 0,961 и 0,958 соответственно, что в совокупности с P -значениями порядка 10^{-10} дает основание считать, что между полученными с помощью предлагаемого метода оценками и реальными характеристиками аккуратности участников есть сильная монотонная зависимость. Вместе с тем в рассматриваемом примере принципиальной разницы между способами расчета характеристики доминирования нет.

Таким образом, предлагаемый метод (с любой из характеристик доминирования) действительно позволяет оценить ожидаемое качество разметки участников, основываясь только на ненадежных и неполных результатах разметки.

Анализ наборов данных с другими значениями параметров (варьирующимся N , a^m) показывает, что наличие такой сильной монотонной зависимости наблюдается при любых значениях параметров. Важным параметром является избыточность при назначении заданий (сколько участников обрабатывают одно изображение). Интуитивно это связано с тем, что чем меньше эта избыточность, тем меньше оказывается данных для сопоставления аккуратности участников, и, наоборот, чем избыточность в назначении изображений выше, тем больше данных для сопоставле-

ния и тем надежней должна оказываться оценка, получаемая таким методом. Действительно, с ростом избыточности коэффициент Спирмена для обоих способов расчета характеристики доминирования увеличивается.

Заключение

В статье предложена модель участника системы совместной разметки и метод оценки ожидаемого качества разметки, получаемой от участников такой системы. Предлагаемый метод основан на получении собственных чисел матрицы характеристики доминирования, построенной определенным образом. Особенность метода заключается в том, что он позволяет получить оценки ожидаемого качества разметки (ранжировать участников по ожидаемому качеству) на основе только ненадежных и неполных результатов разметки, без знания эталонных меток, получение которых часто является трудоемкой задачей. Полученные с помощью предлагаемого метода оценки могут использоваться в алгоритмах обеспечения качества разметки (например, для блокировки недобросовестных участников).

Работа выполнена при финансовой поддержке РФФИ (грант № 16-37-60107).

Литература

1. Wechsler D. Crowdsourcing as a Method of Transdisciplinary Research — Tapping the Full Potential of Participants // *Futures*. 2014. Vol. 60. P. 14–22.
2. Baev V., Sablok G., Minkov I. Next Generation Sequencing Crowd Sourcing at BIOCAMP: What Promises it Holds for us in Future? // *Jour-*

- nal of Computational Science. 2014. Vol. 5. N 3. P. 325–326.
3. Fraternali P., Castelletti A., Soncini-Sessa R., et al. Putting Humans in the Loop: Social Computing for Water Resources Management // *Environmental Modelling and Software*. 2012. Vol. 37. P. 68–77.
4. Nunes A., Galvão T., Cunha J. Urban Public Transport Service Co-creation: Leveraging Passenger's

- Knowledge to Enhance Travel Experience // *Procedia — Social and Behavioral Sciences*. 2014. Vol. 111. P. 577–585.
5. **Brabham D. C., Ribisl K. M., Kirchner T. R., Bernhardt J. M.** Crowdsourcing Applications for Public Health // *American Journal of Preventive Medicine*. N 46(2). P. 179–187.
 6. **Wu H., Sun H., Fang Y., et al.** Combining Machine Learning and Crowdsourcing for Better Understanding Commodity Reviews // *Proc. of the 29th AAAI Conf. on Artificial Intelligence*. 2015. P. 4220–4221.
 7. **Chang J., Amershi S., Kamar E.** Revolt: Collaborative Crowdsourcing for Labeling Machine Learning Datasets // *Proc. of the Conf. on Human Factors in Computing Systems (CHI 2017)*. 2017. P. 2334–2346.
 8. **Lasecki W., Song Y., Kautz H., Bigham J.** Real-time Crowd Labeling for Deployable Activity Recognition // *Proc. of the 2013 Conf. on Computer Supported Cooperative Work*. 2013. P. 1203–1212.
 9. **Irshad H., Oh E.-Y., Schmolze D., et al.** Crowdsourcing Scoring of Immunohistochemistry Images: Evaluating Performance of the Crowd and an Automated Computational Method // *Scientific Reports*. 2017. Vol. 7. Article 43286.
 10. **Leifman G., Swedish T., Roesch K., et al.** Leveraging the Crowd for Annotation of Retinal Images // *Proc. of the 37th Annual Intern. Conf. of the IEEE Engineering in Medicine and Biology Society (EMBC)*. 2015. http://web.media.mit.edu/~gleifman/papers/GSRR_crowd_EMBC15.pdf (дата обращения: 02.05.2018).
 11. **Bernstein A., Klein M., Malone T.** Programming the Global Brain // *Communications of the ACM*. 2012. Vol. 55. Iss. 5. P. 41–43.
 12. **Пономарев А. В.** Разметка изображений массового мероприятия его участниками на основе немонетарного стимулирования // *Информационно-управляющие системы*. 2017. № 3. С. 105–114. doi:10.15217/issn1684-8853.2017.3.105
 13. **Ben-Ami I., Basha T., Avidan S.** Racing Bib Numbers Recognition, *British Machine Vision Conference (BMVC)*. 2012. <http://people.csail.mit.edu/talidekel/papers/RBNR.pdf> (дата обращения: 02.05.2018).
 14. **Roy S., et al.** A New Multi-modal Technique for Bib Number/Text Detection in Natural Images, *Advances in Multimedia Information Processing (PCM 2015)*. 2015. LNCS 9314. P. 483–494.
 15. **Пonomarev A.** Community Photo Tagging: Engagement and Quality Study // *Proc. of the 2017 ACM Web Science Conf. (WebSci'17)*. 2017. P. 409–410.
 16. **Пономарев А. В.** Методы обеспечения качества в системах крауд-вычислений: аналитический обзор // *Тр. СПИИРАН*. 2017. Т. 54. № 5. С. 152–184.
 17. **Zhang J., et al.** Consensus Algorithms for biased Labeling in Crowdsourcing // *Informaion Sciences*. 2017. Vol. 382–383. P. 254–273.
 18. **Matsunaga A., Mast A., Fortes J. a. B. B.** Workforce-efficient Consensus in Crowdsourced Transcription of Biocollections Information // *Future Generation Computer Systems*. 2016. Vol. 56. P. 526–536.
 19. **Dawid A. P., Skene A. M.** Maximum Likelihood Estimation of Observer Error-Rates using the EM Algorithm // *Applied Statistics*. 1979. Vol. 28. N 1. P. 20–28.
 20. **Sheng V. S., Provost F., Ipeirotis P. G.** Get Another Label? Improving Data Quality and Data Mining using Multiple, Noisy Labelers // *Proc. of the 14th ACM SIGKDD Intern. Conf. on Knowledge Discovery and Data Mining (KDD 08)*. 2008. P. 614–622.
 21. **Lou Y., et al.** Use of Ontology Structure and Bayesian Models to Aid the Crowdsourcing of ICD-11 Sanctioning Rules // *Journal of Biomedical Informatics*. 2017. Vol. 68. P. 20–34.
 22. **Dai P., et al.** POMDP-based Control of Workflows for Crowdsourcing // *Artificial Intelligence*. 2013. Vol. 202. P. 52–85.
 23. **Han K., Zhang C., Luo J.** Taming the Uncertainty: Budget Limited Robust Crowdsensing Through Online Learning // *IEEE/ACM Transactions on Networking*. 2016. Vol. 24. N 3. P. 1462–1475.
 24. **Wen Y., et al.** Quality-Driven Auction-Based Incentive Mechanism for Mobile Crowd Sensing // *IEEE Transactions on Vehicular Technology*. 2015. Vol. 64. N 9. P. 4203–4214.
 25. **Shamir L., Diamond D., Wallin J.** Leveraging Pattern Recognition Consistency Estimation for Crowdsourcing Data Analysis // *IEEE Transactions on Human-Machine Systems*. 2016. Vol. 46. N 3. P. 474–480.
 26. **Burmania A., Parthasarathy S., Busso C.** Increasing the Reliability of Crowdsourcing Evaluations using Online Quality Assessment // *IEEE Transactions on Affective Computing*. 2016. Vol. 7. N 4. P. 374–388.
 27. **Гантмахер Ф. Р.** Теория матриц. — М.: Наука, 1966. — 576 с.

UDC 004.9

doi:10.31799/1684-8853-2018-4-45-51

Model and Method for Contributor's Quality Assessment in Community Image Tagging SystemsPonomarev A. V.^a, PhD, Tech., Senior Researcher, ponomarev@iias.spb.su^aSaint-Petersburg Institute for Informatics and Automation of the RAS, 39, 14 Line, V. O., 199178, Saint-Petersburg, Russian Federation

Introduction: Large-scale human-computer systems involving people of various skills and motivation into the information processing process are currently used in a wide spectrum of applications. An acute problem in such systems is assessing the expected quality of each contributor; for example, in order to penalize incompetent or inaccurate ones and to promote diligent ones. **Purpose:** To develop a method of assessing the expected contributor's quality in community tagging systems. This method should only use generally unreliable and incomplete information provided by contributors (with ground truth tags unknown). **Results:** A mathematical model is proposed for community image tagging (including the model of a contributor), along with a method of assessing the expected contributor's quality. The method is based on comparing tag sets provided by different contributors for the same images, being a modification of pairwise comparison method with preference relation replaced by a special domination characteristic. Expected contributors' quality is evaluated as a positive eigenvector of a pairwise domination characteristic matrix. Community tagging simulation has confirmed that the proposed method allows you to adequately estimate the expected quality of community tagging system contributors (provided that the contributors' behavior fits the proposed model). **Practical relevance:** The obtained results can be used in the development of systems based on coordinated efforts of community (primarily, community tagging systems).

Keywords — Crowdsourcing, Crowd Computing, Image Tagging, Image Processing, Collective Intelligence.

Citation: Ponomarev A. V. Model and Method for Contributor's Quality Assessment in Community Image Tagging Systems. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2018, no. 4, pp. 45–51 (In Russian). doi:10.31799/1684-8853-2018-4-45-51

References

- Wechsler D. Crowdsourcing as a Method of Transdisciplinary Research — Tapping the Full Potential of Participants. *Futures*, 2014, vol. 60, pp. 14–22.
- Baev V., Sablok G., Minkov I. Next Generation Sequencing Crowd Sourcing at BIOCOMP: What Promises it Holds for us in Future? *Journal of Computational Science*, 2014, vol. 5, no. 3, pp. 325–326.
- Fraternali P., Castelletti A., Soncini-Sessa R., et al. Putting Humans in the Loop: Social Computing for Water Resources Management. *Environmental Modelling and Software*, 2012, vol. 37, pp. 68–77.
- Nunes A., Galvão T., Cunha J. Urban Public Transport Service Co-creation: Leveraging Passenger's Knowledge to Enhance Travel Experience. *Procedia — Social and Behavioral Sciences*, 2014, vol. 111, pp. 577–585.
- Brabham D. C., Ribisl K. M., Kirchner T. R., & Bernhardt J. M. Crowdsourcing Applications for Public Health. *American Journal of Preventive Medicine*, 2014, no. 46(2), pp. 179–187.
- Wu H., Sun H., Fang Y., et al. Combining Machine Learning and Crowdsourcing for Better Understanding Commodity Reviews. *Proc. of the 29th AAAI Conf. on Artificial Intelligence*, 2015, pp. 4220–4221.
- Chang J., Amershi S., Kamar E. Revolt: Collaborative Crowdsourcing for Labeling Machine Learning Datasets. *Proc. of the Conf. on Human Factors in Computing Systems (CHI 2017)*, 2017, pp. 2334–2346.
- Lasecki W., Song Y., Kautz H., Bigham J. Real-time Crowd Labeling for Deployable Activity Recognition. *Proc. of the 2013 Conf. on Computer Supported Cooperative Work*, 2013, pp. 1203–1212.
- Irshad H., Oh E.-Y., Schmolze D., et al. Crowdsourcing Scoring of Immunohistochemistry Images: Evaluating Performance of the Crowd and an Automated Computational Method. *Scientific Reports*, 2017, vol. 7, article 43286.
- Leifman G., Swedish T., Roesch K., et al. Leveraging the Crowd for Annotation of Retinal Images. *Proc. of the 37th Annual Intern. Conf. of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2015. Available at: http://web.media.mit.edu/~gleifman/papers/GSRR_crowd_EMBC15.pdf (accessed 2 May 2018).
- Bernstein A., Klein M., Malone T. Programming the Global Brain. *Communications of the ACM*, 2012, vol. 55, iss. 5, pp. 41–43.
- Ponomarev A. Tagging Public Event Images based on Non-Monetary Incentives. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2017, no. 3, pp. 105–114 (In Russian). doi:10.15217/issn1684-8853.2017.3.105
- Ben-Ami I., Basha T., Avidan S. Racing Bib Numbers Recognition. *British Machine Vision Conf. (BMVC)*, 2012. Available at: <http://people.csail.mit.edu/talidekel/papers/RBNR.pdf> (accessed 2 May 2018).
- Roy S., et al. A New Multi-modal Technique for Bib Number/Text Detection in Natural Images. *Advances in Multimedia Information Processing — PCM 2015*, 2015, LNCS 9314, pp. 483–494.
- Ponomarev A. Community Photo Tagging: Engagement and Quality Study. *Proc. of the 2017 ACM Web Science Conf.*, 2017, pp. 409–410.
- Ponomarev A. Quality Control Methods in Crowd Computing: Literature Review. *Trudy SPIIRAN* [SPIIRAS Proceedings], 2017, vol. 54, no. 5, pp. 152–184 (In Russian).
- Zhang J., et al. Consensus Algorithms for biased Labeling in Crowdsourcing. *Informaion Sciences*, 2017, vol. 382–383, pp. 254–273.
- Matsunaga A., Mast A., Fortes J. a. B. B. Workforce-efficient Consensus in Crowdsourced Transcription of Biocollections Information. *Future Generation Computer Systems*, 2016, vol. 56, pp. 526–536.
- Dawid A. P., Skene A. M. Maximum Likelihood Estimation of Observer Error-Rates using the EM Algorithm. *Applied Statistics*, 1979, vol. 28, no. 1, pp. 20–28.
- Sheng V. S., Provost F., Ipeirotis P. G. Get Another Label? Improving Data Quality and Data Mining using Multiple, Noisy Labelers. *Proc. of the 14th ACM SIGKDD Intern. Conf. on Knowledge Discovery and Data Mining (KDD 08)*, 2008, pp. 614–622.
- Lou Y., et al. Use of Ontology Structure and Bayesian Models to Aid the Crowdsourcing of ICD-11 Sanctioning Rules. *Journal of Biomedical Informatics*, 2017, vol. 68, pp. 20–34.
- Dai P., et al. POMDP-based Control of Workflows for Crowdsourcing. *Artificial Intelligence*, 2013, vol. 202, pp. 52–85.
- Han K., Zhang C., Luo J. Taming the Uncertainty: Budget Limited Robust Crowdsensing Through Online Learning. *IEEE/ACM Transactions on Networking*, 2016, vol. 24, no. 3, pp. 1462–1475.
- Wen Y., et al. Quality-Driven Auction-Based Incentive Mechanism for Mobile Crowd Sensing. *IEEE Transactions on Vehicular Technology*, 2015, vol. 64, no. 9, pp. 4203–4214.
- Shamir L., Diamond D., Wallin J. Leveraging Pattern Recognition Consistency Estimation for Crowdsourcing Data Analysis. *IEEE Transactions on Human-Machine Systems*, 2016, vol. 46, no. 3, pp. 474–480.
- Burmania A., Parthasarathy S., Busso C. Increasing the Reliability of Crowdsourcing Evaluations using Online Quality Assessment. *IEEE Transactions on Affective Computing*, 2016, vol. 7, no. 4, pp. 374–388.
- Gantmaher F. R. *Teoriya matric* [Matrix Theory]. Moscow, Nauka Publ., 1966. 576 p. (In Russian).