

А.В. ПОНОМАРЕВ  
**МЕТОДЫ ОБЕСПЕЧЕНИЯ КАЧЕСТВА РЕЗУЛЬТАТОВ В  
СИСТЕМАХ КРАУД-ВЫЧИСЛЕНИЙ: АНАЛИТИЧЕСКИЙ  
ОБЗОР**

*Пономарев А.В. Методы обеспечения качества результатов в системах крауд-вычислений: аналитический обзор.*

**Аннотация.** В настоящее время все большее распространение получают крауд-вычисления, то есть привлечение к задачам обработки информации широкого круга людей, взаимодействующих посредством инфокоммуникационных технологий. Тем не менее практическое применение этой технологии в значительной мере сдерживается неопределенностью качества получаемых результатов. В этих условиях задача систематизации сведений об используемых на данный момент методах обеспечения качества и идентификации перспективных направлений их развития является особенно актуальной. В статье обсуждаются результаты систематического обзора журнальных публикаций полнотекстовых баз ScienceDirect и IEEE Xplore, вышедших после 2012 года. Выделены наиболее распространенные на данный момент направления в обеспечении качества, используемые модели и принимаемые допущения, обозначены границы применимости методов. Отмечено, что наибольшее распространение получили методы, основанные на согласовании оценок, полученных от разных участников, и методы, основанные на применении теоретико-игровых моделей.

**Ключевые слова:** крауд-вычисления, человеко-машинные вычисления, краудсорсинг, социальные вычисления, человеко-машинные системы, человеческие факторы, систематический обзор литературы.

**1. Введение.** Крауд-вычисления (англ. *crowd computing*), или масштабные человеко-машинные вычисления в настоящее время широко применяются для выполнения различных задач, для которых не существует надежных способов автоматического решения, но которые являются естественными для человека и в которых человек может быть эффективен (например, разметка и аннотирование изображений и аудиозаписей, обработка сложных данных и данных с неопределенностью). Слово «вычисления» в составе этого термина трактуется максимально широко, не ограничиваясь рамками математических вычислений и неся скорее значение «обработка информации». Так, одно из популярных применений крауд-вычислений — создание размеченных эталонных наборов для задач машинного обучения (например, [1]). В этом случае человек-«вычислитель» обрабатывает какой-то объект (как правило, изображение, аудио- или видеофрагмент), и результатом обработки (или «вычислений») является некоторая метка, связанная с содержанием объекта и отвечающая цели создания набора.

Использование крауд-вычислений сопряжено с рядом принципиальных ограничений [2], среди которых проблема качества

результатов, получаемых таким образом, является наиболее острой и во многом именно она ограничивает применение крауд-вычислений.

Причиной этого является как общая склонность к ошибкам человека, включаемого в процесс обработки информации (даже при добросовестном выполнении заданий), так и различные виды сознательных нарушений инструкций участником для получения максимальной выгоды при минимизации затрачиваемых усилий (например, [3]).

Под качеством в широком смысле понимается соответствие результата выполнения некоторой задачи (сбора или обработки информации) с помощью крауд-вычислений и «истинного» результата этой задачи. Роль последнего на практике, как правило, выполняет результат, полученный заинтересованным в выполнении задачи компетентным «экспертом». Конкретные способы оценки качества в значительной мере зависят от области применения. Например, в задачах разметки результатом крауд-вычислений является множество меток, присвоенных объектам участниками, и для оценки качества этого множества часто используются меры точности (precision) и полноты (recall), позаимствованные из информационного поиска, а в роли эталонного множества выступает множество меток, полученных от эксперта.

Цель данного исследования — выявление основных подходов и методов, используемых в настоящее время для управления качеством информации, получаемой с помощью краудсорсинга (управления качеством решения задач с применением краудсорсинга), а также идентификация актуальных направлений исследований в этой области.

Для достижения поставленной цели был проведен отбор современных научных публикаций, касающихся вопросов обеспечения качества, а затем на основе анализа этих публикаций выявлены основные методы и модели, определены их особенности и типовые области применения.

**2. Методика исследования.** В основу исследования положена методика проведения систематических обзоров литературы, предложенная Б. Китченхем [4, 5]. Одним из важнейших достоинств этой методики является воспроизводимость результатов за счет точной спецификации используемых источников информации и поисковых запросов, а также фиксации всех промежуточных результатов в протоколе исследования. Вместе с тем строгое применение данной методики для анализа обширных областей знаний связано с рядом технических трудностей — в первую очередь, это управление размером выборки исследуемых публикаций и учет их значимости. Для преодоления этих трудностей методика была расширена.

Систематический обзор был применен к относительно небольшому множеству современных публикаций, затем множество исследуемых публикаций было расширено за счет приставейных списков литературы отобранных публикаций.

Исследование, результаты которого изложены в данной статье, состояло из двух этапов:

1) Систематический обзор современной литературы по методике Б. Китченхем, ставящий целью выявление актуальных методов и подходов к обеспечению качества в системах крауд-вычислений.

2) Анализ источников, цитируемых в отобранных статьях, для идентификации наиболее значимых исторических моделей и методов (к которым, в частности, предлагаются усовершенствования в публикациях, отобранных на этапе 1). Эта информация дополняет современную проблематику, выявленную на этапе 1, и позволяет раскрыть историческую перспективу и логику развития данной области исследований.

В качестве источника публикаций для систематического обзора были использованы две полнотекстовые библиографические базы: ScienceDirect и IEEE Xplore. В каждой из баз был осуществлен поиск статей в периодических изданиях (журналах), опубликованных за последние 5 лет (с 2012 года), и удовлетворяющих следующему поисковому запросу: «quality AND ("crowd computing" OR crowdsourcing OR "crowd-sourcing" OR "crowd sourcing" OR "social computing" OR "social computations" OR "human computations")» или аналогичному (с учетом поддерживаемого библиографической базой поискового языка). Таким образом, дальнейшему анализу подвергались статьи, название, аннотация или ключевые слова которых содержали слово quality (качество) и один из терминов, которыми наиболее часто описывается процесс масштабных человеко-машинных вычислений, производимых с помощью сети Интернет. Ограничение поиска статьями в периодических изданиях (журналах) вызвано, во-первых, тем, что результаты, опубликованные в таких изданиях, как правило, прошли более серьезную апробацию и строгий отбор, во-вторых, тем, что это позволяет снизить общий объем первичного материала для ознакомления (сконцентрировавшись на — предположительно — наиболее апробированном его подмножестве). В результате выполнения запросов из базы ScienceDirect было отобрано 124 статьи, а из IEEE Xplore — 79.

Затем в ходе скрининга аннотаций полученных 203 статей были отобраны статьи, удовлетворяющие хотя бы одному из следующих условий: а) статья посвящена исследованию того или иного способа обеспечения качества при использовании краудсорсинга (например [1, 6]); б) статья посвящена описанию способа применения краудсорсинга

для решения практической задачи, но в аннотации авторы подчеркивают важность обеспечения качества и заявляют об использовании того или иного подхода или метода (независимо от степени оригинальности) — например [7, 8]. Вторичные публикации (обзоры) были отнесены в отдельную категорию. В результате для дальнейшего исследования была отобрана 41 статья, из которых 3 обзорных.

При выполнении второго этапа исследуемая литература уже не ограничивалась определенными библиографическими базами и характером публикации (принимались во внимание и не только публикации в периодических изданиях). Решающую роль здесь играла частота (и семантика) ссылок из статей, отобранных в первом этапе.

Таким образом, используемая методика позволяет одновременно получить объективный «срез» публикаций на заданную тему в авторитетных периодических изданиях и выявить базовые методы и наиболее влиятельные публикации.

**3. Количественный анализ и типизация.** Для определения фазы, в которой находятся исследования в данной области, была проанализирована динамика количества публикаций, удовлетворяющих сформулированному запросу (рисунок 1). Приведенный график показывает, что увеличение количества публикаций происходит нарастающими темпами, то есть проблема переживает фазу активного исследования, сопровождающегося значительным энтузиазмом по поводу потенциала систем, основанных на краудсорсинге и крауд-вычислениях. Заметим также, что учет общего роста количества публикаций в выбранных полнотекстовых базах со временем не вносит существенных изменений в представленную закономерность.

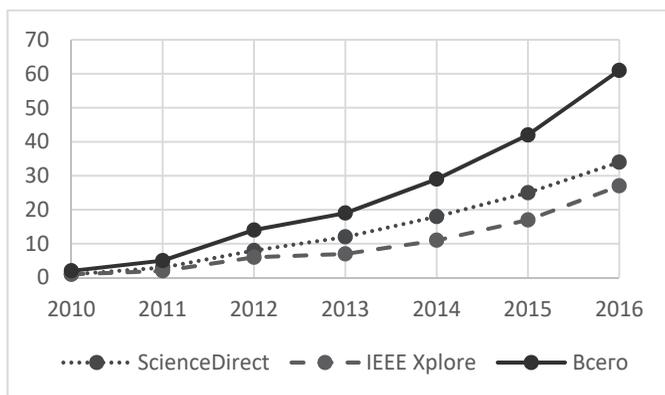


Рис. 1. Количество статей, посвященных исследуемой тематике, по годам (в полнотекстовых библиографических базах ScienceDirect и IEEE Xplore)

**3.1. Классификация.** В ходе детального анализа отобранных публикаций выделено 6 основных направлений, по которым могут быть сгруппированы методы обеспечения качества: методы согласования (consensus), методы проектирования потока работ, методы централизованного назначения работ, теоретико-игровые методы, методы, основанные на учете свойств заданий, и методы, основанные на анализе действий пользователя и воздействии на него. Помимо этого, в отдельные категории отнесены обзорные статьи и статьи, которые были признаны нерелевантными после анализа полного текста. Количественное распределение работ по категориям приведено в таблице 1.

Таблица 1. Классификация отобранных работ по направлениям

Направление	Ссылки	Количество
Методы согласования	[6, 7, 9-17]	11
Методы проектирования потока работ	[18-21]	4
Методы централизованного назначения	[8, 22-25]	5
Теоретико-игровые методы	[26-32]	7
Методы, основанные на учете свойств заданий	[1, 33, 34]	3
Методы, основанные на анализе действий пользователя и воздействии на него	[35-37]	3
Обзорные статьи	[38-40]	3
Нерелевантные (исключенные)	[41-45]	5

**3.2. Обзорные статьи.** В ряде обзорных статей неизбежно затрагивается вопрос обеспечения качества. В сформированную выборку вошли следующие обзорные статьи.

Статья [39] содержит большое количество вводной информации по многим аспектам, касающимся организации краудсорсинга, а также богатую библиографию, однако непосредственно вопросам обеспечения качества уделено не так много внимания.

Обзорная статья [40] посвящена широкому кругу проблем, связанных с данными, получаемыми с помощью краудсорсинга. В частности, выделяется три проблемы, связанные с этим: управление качеством, управление стоимостью и управление латентностью (временем). Приводится достаточно подробный обзор различных методов организации краудсорсинга, а также применений его в области обработки данных, однако рассуждения об управлении качеством в основном ограничены методами согласования, что несколько уже, чем в настоящей статье.

Обзорная статья [38] содержит обзор основных методов обеспечения качества. В частности, предлагает лаконичную платформу для классификации различных методов обеспечения качества, однако некоторые методы, набирающие популярность в последние годы, остались неосвещенными (в первую очередь, это касается теоретико-игровых методов и методов централизованного назначения).

Таким образом, настоящая статья, будучи ориентированной на обзор современных подходов к обеспечению качества и представляющая результаты обобщения четко определенного множества первичных работ, отличается от перечисленных статей, дополняет их и, возможно, наряду с ними послужит распространению крауд-вычислений и дальнейшему развитию исследований в этой области.

**3.3. Нерелевантные статьи.** Ряд статей, отобранных в результате скрининга аннотаций, оказались не вполне соответствующими цели исследования. В первую очередь, это статьи по пространственному краудсорсингу, в которых поднимается вопрос качества информации (Quality of Information, QoI), однако интерпретируется это качество просто как выполнение требований по покрытию, в определенном смысле, полнота информации: [41, 42, 44].

**4. Направления в обеспечении качества.** При применении крауд-вычислений типовым является следующий сценарий:

- разработка задания;
- распределение заданий между участниками (исполнителями);
- выполнение заданий участниками;
- обработка результатов.

Поскольку методы обеспечения качества имеют первостепенную важность в практическом применении таких систем, существующие (в том числе и найденные в процессе обзора литературы) методы охватывают все этапы этого сценария. Применимость методов различных направлений к этапам типового сценария отображена в таблице 2.

Таблица 2. Охват различных направлений в обеспечении качества

Направление	Разработка задания	Распределение заданий	Выполнение	Обработка результатов
Методы согласования				+
Методы проектирования потока работ		+	+	+
Методы централизованного назначения		+		
Теоретико-игровые методы	+	+	+	+
Методы, основанные на учете свойств заданий			+	+
Методы, основанные на анализе действий пользователя и воздействия на него	+		+	+

Помимо этого, ключевую роль при обсуждении методов обеспечения качества играют: а) набор допущений, которые делаются относительно задач, исполнителей и имеющейся информации об исполнителе; б) дополнительные характеристики (помимо качества) процесса, которые необходимо учесть (скорость решения задачи, стоимость и пр.).

Оставшаяся часть данного раздела построена следующим образом. Отдельный подраздел выделен под каждое из упомянутых направлений; в начале подраздела дается краткая характеристика направления, основные идеи, которые лежат в его основе (для пояснения этих базовых идей используются источники, отобранные на втором этапе обзора литературы — часто основополагающие и особенно цитируемые публикации), затем приводится краткая характеристика всех современных работ по соответствующему направлению, удовлетворяющих условиям систематического обзора литературы (этап 1).

**4.1. Методы согласования (результатов, полученных от разных исполнителей).** Данные методы получили широкое распространение в тех вариантах организации краудсорсинга, где *недоступна априорная информация об исполнителях* (например, об ожидаемом качестве выполнения заданий исполнителем) и о заданиях (например, их трудность). В этом смысле алгоритмы согласования являются наиболее универсальными, потому что накладывают минимальные ограничения на содержимое профиля исполнителя и могут работать с анонимными исполнителями. Общий принцип работы таких методов заключается в том, что каждое задание выполняется несколькими участниками системы крауд-вычислений, затем полученные результаты обобщаются (согласуются) для получения итогового результата.

Простейшим алгоритмом такого обобщения является мажоритарное голосование (majority voting, MV) [46], которое при определенных допущениях позволяет получать результаты, качество которых является достаточным для многих прикладных задач [6].

Недостаток мажоритарного голосования в том, что при агрегации данных, полученных от разных участников, значимость этих фрагментов считается одинаковой, что полностью игнорирует возможные различия в уровне компетентности и добросовестности участников. Более сложные алгоритмы пытаются преодолеть этот недостаток. Один из основных алгоритмов такого плана — алгоритм DS, предложенный в [47], являющийся одним из наиболее часто упоминаемых в статьях, посвященных методам поиска консенсуса, и

очень часто выступающий в качестве базового при оценке предлагаемых методов (например, в работах [6, 10, 16]).

Методы согласования естественным образом разбиваются на две группы — методы, основанные на EM-алгоритме (expectation minimization), и методы, не основанные на EM-алгоритме. В первую группу попадают, например, DS, GLAD [46], RY [48], KOS [49], Spectral DS [50], ZenCrowd [51], а во вторую — мажоритарное голосование, AWMV [6].

EM-алгоритм — широко распространенный алгоритм математической статистики для нахождения оценок максимального правдоподобия параметров вероятностных моделей [52, 53]. Все алгоритмы согласования (поиска консенсуса), основанные на EM-алгоритме, одновременно оптимизируют параметры модели и вычисляют обобщенные ответы для заданий (в идеале, приближающиеся к «истинным»). Алгоритмы согласования, относящиеся к семейству EM, в сущности, различаются набором параметров модели, которые определяют ответ на задание, получаемый от участника. Так, например, в DS моделью является матрица ошибок каждого участника, GLAD — уровни компетентности участников и уровни сложности заданий, через взаимодействие которых выражаются обобщенные оценки и так далее.

Формально задача поиска консенсуса формулируется следующим образом [6]. Задан набор примеров  $E = \{e_i\}_{i=1}^I$ , где каждый пример  $e_i = (x_i, y_i)$  состоит из набора значений признаков  $x_i$  и истинной метки  $y_i$ . Множество участников системы будем обозначать  $U = \{u_j\}_{j=1}^J$ . Каждая метка, полученная от участника, принадлежит множеству классов  $C = \{c_k\}_{k=1}^K$ . Для удобства индексы элементов этих множеств используются как идентификаторы соответствующих объектов, то есть можно говорить о том, что участник  $i$  отнес пример  $j$  к классу  $k$ . Чаще всего рассматривается (наиболее простая) задача с двумя классами; в этом случае класс  $c_1$  ( $k=1$ ) соответствует положительным («+») примерам, а класс  $c_2$  ( $k=2$ ) — отрицательным («-»). Каждому примеру  $i$  соответствует множество меток  $I_i = \{I_{ij}\}_{j=1}^J$ , элемент  $I_{ij}$  которого получен от участника  $j$ . Все метки примеров в наборе данных формируют матрицу аннотаций  $L = \{I_i\}_{i=1}^I$ ,  $I_{ij} \in \{c_1, 0, c_2\}$ , где 0 означает, что участник не предоставил метки для данного примера. Участник  $j$  характеризуется также

матрицей  $N^{(j)} = \{n_{ik}^j\}, 1 \leq i \leq I, 1 \leq k \leq K$ , каждый элемент которой равен количеству раз, когда участник  $i$  отнес пример  $j$  к классу  $k$ . На практике, как правило,  $n_{ik}^j \in \{0, 1\}$ . Известны также априорные вероятности каждого из классов.

Задача состоит в том, чтобы найти метки  $\hat{y}_i$ , минимизирующие функцию ошибки:

$$R = \frac{1}{I} \sum_{i=1}^I \mathbf{I}(\hat{y}_i \neq y_i)$$

при заданных  $L$  и  $y \in (c_1, c_2)$ . Здесь  $\mathbf{I}$  — индикаторная функция.

Алгоритм DS решает эту задачу следующим образом. Пусть каждый участник  $j$  системы описывается матрицей ошибок классификации (confusion matrix)  $\pi_{kl}^{(j)}$ , задающей вероятность того, что данный участник укажет класс  $l$  для примера, истинным классом которого является  $k$ . На E-шаге по известной матрице аннотаций  $L$  алгоритм оценивает вероятности того, что пример  $i$  принадлежит классу  $k$  (для каждого класса) в соответствии со следующим уравнением:

$$P(\hat{y}_i = c_k | L) = \frac{\prod_{j=1}^J \prod_{l=1}^L (\pi_{kl}^{(j)})^{n_{il}^j} P(c_k)}{\sum_{q=1}^K \prod_{j=1}^J \prod_{l=1}^L (\pi_{ql}^{(j)})^{n_{il}^j} P(c_q)}$$

где  $P(c_k)$  и  $P(c_q)$  — априорные вероятности принадлежности примера классам  $c_k$  и  $c_q$  соответственно, вычисленные на M-шаге алгоритма. На M-шаге происходит уточнение значений матрицы ошибок классификации и априорных вероятностей классов в соответствии со следующими соотношениями:

$$\hat{\pi}_{kl}^{(j)} = \frac{\sum_{i=1}^I \mathbf{I}(\hat{y}_i = c_k) n_{il}^{(j)}}{\sum_{l=1}^K \sum_{i=1}^I \mathbf{I}(\hat{y}_i = c_k) n_{il}^{(j)}},$$

$$\hat{P}(c_k) = \frac{1}{I} \sum_{i=1}^I \mathbf{I}(\hat{y}_i = c_k).$$

Поскольку истинные классы примеров неизвестны, вместо  $\mathbf{I}(\hat{y}_i = c_k)$  используется математическое ожидание этой величины, то есть  $P(\hat{y}_i = c_k | L)$ .

На вход алгоритма подаются начальные оценки «правильных» меток для примеров. На основе этого выполняется М-шаг, в ходе которого вычисляются матрица ошибок и априорные вероятности классов, далее выполняется Е-шаг, на котором вычисляются новые оценки вероятностей принадлежности примеров каждому из классов, и так до достижения сходимости.

Большинство методов согласования так или иначе обладает следующими ограничениями. Во-первых, тип ответа на задание. В большинстве публикаций рассматриваются задания, предполагающие бинарные ответы, то есть суть задания сводится к бинарной классификации какого-то объекта. Как правило, бинарная классификация может быть расширена на выбор одного из множества ответов, однако это усложняет и без того достаточно сложные модели. Во-вторых, значительная вычислительная сложность моделей (особенно основанных на EM-алгоритме). В-третьих, для работы алгоритмов, основанных на EM, необходимо, чтобы каждое задание было выполнено значительным количеством участников (около 10), что может быть связано с существенными финансовыми затратами, если задания выполняются на коммерческой основе. Наконец, при практической реализации могут возникнуть трудности с определением начальных значений параметров для EM-алгоритма [16].

Тем не менее, несмотря на определенные ограничения, наличие строгого математического базиса привело к тому, что методы обеспечения качества в крауд-вычислениях, основанные на согласовании мнений различных участников, оказываются, пожалуй, наиболее проработанными и получают значительное внимание исследователей. Так, среди 41 публикации, вошедшей в систематический обзор, 11 посвящены различным методам согласования (таблица 1).

Следует также упомянуть попытки стандартизованного сравнения различных алгоритмов согласования, например, SQUARE [54] или BATC [55].

Современные работы, вошедшие в систематический обзор, посвящены развитию методов согласования по следующим направлениям.

Важным направлением развития методов согласования является разработка специализированных методов, учитывающих особенности того или иного класса задач (и множества меток). Так, в ряде работ рассматривается ситуация, когда между метками, которые являются ожидаемым результатом выполнения задания участником, существуют определенные логические связи. Несмотря на то что на практике такая

ситуация встречается достаточно часто, хорошо разработанных и принятых научным сообществом методов агрегации меток для нее пока не существует. Разработке таких методов посвящена работа [10], авторы которой предлагают расширение алгоритма DS и исследуют разные модели представления отношений между метками (включая байесовскую сеть как компактное представление совместного распределения). В статье [13] предлагается вероятностная модель разметки для иерархической классификации, то есть для ситуации, когда метками, которые участники присваивают объектам, являются классы, организованные в иерархию (классификация книг, товаров). Адаптация и разработка методов согласования для заданий, выходящих за рамки простых бинарных схем, рассматривается также в работах [7, 11, 15]. В [11] предлагается метод согласования ответов в непрерывной шкале, в [15] речь идет о методе согласования для ответов в рамках порядковой шкалы, а в [7] — о методах согласования текстовых данных.

Особой темой исследований в области методов поиска консенсуса является разработка способов учета смещения (bias) в определении меток [6, 17], то есть, ситуации, когда несмотря на то, что участники добросовестно выполняют свою задачу, восприятие размечаемых объектов у разных групп пользователей различается. Например, у неспециалистов (коими обычно являются участники систем краудсорсинга) может быть склонность классифицировать объекты не так, как у специалистов.

Еще одним направлением является учет различных моделей поведения пользователей при формировании обобщенной оценки. Так, в работе [11] предлагается метод выявления согласованной оценки для непрерывных измерений при отсутствии золотого стандарта и векторов свойств объектов. Авторы статьи рассматривают различные стили поведения исполнителей (спамер, смещенный, злоумышленник и пр.) и формализуют их в виде байесовских моделей. В статье [15] предлагается способ выявления мусорных (спамерских) результатов. Выявление спамеров базируется на наблюдении, что популярным методом спама является генерация случайных результатов и генерация одних и тех же результатов (для всех заданий исполнителя).

Предлагаются также альтернативные (по отношению к популярному EM-алгоритму) способы поиска консенсуса. Так, в статье [16] предлагается алгоритм получения обобщенных меток на основе кластеризации. В задаче разметки, в которой каждому объекту может быть присвоено несколько меток (несколько классов), вероятность принадлежности каждому из классов, оцененная на основе собранных меток с помощью метода максимального правдоподобия,

рассматривается как вектор признаков объекта. Объекты кластеризуются по сходству этих векторов и итоговая метка для всех объектов, попавших в кластер, определяется как признак с наибольшей общей суммой. В статье [14] предложен оригинальный подход к обеспечению качества краудсорсинга, заключающийся в применении помехоустойчивого кодирования.

В статье [12] решается интересная задача определения оптимального количества эталонных значений (ground truths) и распределения их среди исполнителей для получения максимального качества разметки (в рамках применения краудсорсинга для разметки набора данных). Как и в большинстве алгоритмов, основанных на EM, предполагается, что есть существенное пересечение множеств заданий, выполненных различными участниками. Необычной особенностью предлагаемого авторами подхода является также оценивание доверительного интервала точности исполнителей.

Таким образом, на основе проведенного анализа можно сделать следующие выводы. Ключевыми особенностями методов согласования являются:

- использование вероятностной модели участника процесса крауд-вычислений (это может быть просто вероятность правильного ответа, может быть матрица вероятностей ошибок и пр.);
- возможность применения в условиях минимальной информации об участниках. В то же время многие методы (особенно основанные на EM) позволяют учесть имеющуюся информацию);
- необходимость относительно большого пересечения множеств заданий, обработанных разными участниками.

Текущими направлениями развития методов согласования являются:

- разработка методов согласования для заданий с множеством вариантов ответа, в особенности, когда элементы этого множества взаимосвязаны;
- оценка доверительного интервала параметров модели (например, ожидаемой точности участника);
- разработка методов, обладающих высокой вычислительной эффективностью и обеспечивающих хорошее качество согласования;
- учет субъективных характеристик участников, разработка более точных моделей поведения участников, учитывающих их субъективные характеристики (bias).

**4.2. Методы проектирования потока работ.** Все рассмотренные ранее модели обеспечения качества, основанные на согласовании ответов, полученных от разных участников, предполагали простейший сценарий взаимодействия человека и системы крауд-

вычислений — разным участникам назначаются однотипные задания, участники дают ответы (выполняют задания), полученные ответы обобщаются с помощью того или иного математического алгоритма, чтобы получить истинные (наиболее вероятные) ответы для заданий. Причем для каждого задания, как правило, осуществляется сбор фиксированного количества ответов от участников. Поток работ в крауд-вычислениях может быть и гораздо более сложным. Он может предполагать разные типы заданий, служащие для одной цели (например, генерация набора меток и проверка релевантности конкретной метки), учитывать разные компетенции участников и, главное, быть адаптивным, динамически подстраиваться под текущую ситуацию — оценивать нехватку информации и имеющееся множество ресурсов (исполнителей), формируя на основе этой оценки необходимые задания, стремясь тем самым сократить общее количество работы, которая должна быть выполнена для получения результата определенного качества.

В ряде вошедших в выборку работ рассматриваются именно такие методы обоснованного формирования (в том числе динамического) сложных потоков работ, учитывающих различные параметры.

Научные исследования здесь можно условно разбить на две ветви. Во-первых, это разработка математических (в первую очередь, аналитических) моделей существующих и широко распространенных механизмов обеспечения качества — мажоритарного голосования, контрольных групп и так далее, и дальнейшее использование этих математических моделей для рационального формирования потоков работ. К этой ветви относится, например, статья [19], в которой производится сравнение мажоритарного голосования и использования контрольной группы (нанимаемой из состава участников). Авторами построена аналитическая модель, позволяющая увязать вероятность правильного выполнения задачи и ее стоимость.

Ко второй ветви относятся работы, в которых предлагаются оригинальные математические модели и методы формирования потока работ. Так, в статье [20] исследуются методы нахождения наилучших альтернатив из заданного набора (top K). Задача формулируется с помощью трех ключевых факторов: латентность (задержка), денежная стоимость и качество ответов. Статья [21] посвящена использованию краудсорсинга для создания набора правил, позволяющих формировать осмысленные правила сочетания характеристик для описания заболеваний в соответствии со схемой кодирования заболеваний ICD 11. В частности, использовались моделирование на основе байесовской сети и обучение с подкреплением. Особенность работы

заключается в том, что авторы предлагают динамическую модель потока работ; на основе байесовской сети определяется вероятность того или иного ответа, и если вероятность оказывается выше пороговой, то задание больше не назначается. В статье [18] предлагается разработка агентов, основанных на обучении байесовских сетей в сочетании с частично наблюдаемыми марковскими процессами принятия решений (POMDPs) для поиска компромисса между стоимостью и качеством потоков работ, выполняемых с помощью краудсорсинга. Основная идея работы также состоит в динамической адаптации потока работ в зависимости от получаемых результатов.

В качестве оценки таких методов помимо практических экспериментов используется и имитационное моделирование [20].

Подводя краткий итог по этой группе методов, следует отметить следующее. Методы проектирования потока работ можно рассматривать как своего рода развитие методов обобщения. Их общей чертой является, например, применение вероятностных моделей для описания поведения и качества результатов, получаемых от отдельных участников. Однако методы динамического проектирования позволяют рассматривать и другие важные характеристики процесса (в том числе стоимость), позволяя получать результаты определенного качества потенциально эффективнее, чем методы согласования.

**4.3. Методы централизованного назначения.** Одной из достаточно типовых постановок является выбор распределения заданий между исполнителями, максимизирующий ожидаемое качество результата. При этом неизбежно возникает вопрос о том, как это ожидаемое качество оценивать. Популярным ответом на этот вопрос является долговременная характеристика пользователя, обычно называемая «репутацией». При этом считается, что ожидаемое качество результата, получаемого от участника, не убывает с ростом его репутации. Существует множество моделей для вычисления репутации [56-58], а также работ по противодействию попыткам «обмана» механизма уточнения репутации (например, [59, 60]).

Наибольшее распространение методы централизованного назначения получили в контексте некоторых прикладных разновидностей краудсорсинга, в частности пространственного краудсорсинга и краудсенсинга, то есть получения информации о состоянии физической среды с использованием мобильных устройств участников ([22, 23]). Вероятно, это связано с тем, что в подобных приложениях при распределении заданий традиционно учитывается целый ряд факторов (покрытие необходимой области пространства, расходование энергии, затраты на передачу трафика), качество

получаемой информации становится еще одним из критериев, вписываемых в структуру оптимизационной задачи.

Базовая модель такого типа записывается следующим образом. Пусть  $r_{ij}$  — ожидаемое качество выполнения работы  $i$  исполнителем  $j$  (репутация),  $c_{ij}$  — стоимость выполнения работы  $i$  исполнителем  $j$ ,  $B$  — бюджет рекрутера,  $L_j$  — количество работ, которое может выполнить исполнитель  $j$ , а  $x_{ij}$  — бинарная переменная, принимающая значение 1 в случае назначения работы  $i$  участнику  $j$  и 0 в противном случае. Тогда:

$$\begin{aligned} & \max \sum_{i=1}^N \sum_{j=1}^M x_{ij} r_{ij} \\ & \sum_{i=1}^N x_{ij} \leq L_j, \quad j \in \{1, \dots, M\} \\ & \sum_{i=1}^N \sum_{j=1}^M x_{ij} c_{ij} \leq B. \end{aligned}$$

То есть, задача заключается в том, чтобы найти такое распределение работ по исполнителям (участникам системы крауд-вычислений), чтобы максимизировать ожидаемое качество (считая его для простоты линейной функцией), при условии ограничений на загрузку каждого исполнителя и бюджет рекрутера.

При рассмотрении централизованного назначения математические модели и, соответственно, методы обеспечения качества, основанные на использовании этих моделей, можно разделить на две группы. В первую группу входят методы, в которых процесс обновления репутации рассматривается независимо от процесса распределения задач (например, [24]), то есть при распределении задач ожидаемое качество результатов, получаемого от каждого из участников, считается известным. Во вторую группу входят более общие методы, в которых предполагаемое качество каждого участника считается неизвестным, и процесс организуется таким образом, чтобы сначала оценить ожидаемое качество, а затем распределить задания между участниками на основании этих оценок (например, [8, 25]).

К первой группе относится работа [24], в которой обеспечение качества основывается на учете репутации, то есть характеристики исполнителя, отражающей его «надежность» при выполнении заданий. Причем, поскольку данная статья ориентирована на разработку методов, используемых в системах пространственного краудсорсинга, то в ней решается задача оптимизации распределения заданий с учетом

как репутации, так и пространственного положения пользователей. В частности, задача моделируется как задача о рюкзаке (вариант с несколькими рюкзаками — Multiple Choice Knapsack Problem). В силу сложности данной задачи для ее решения применяют различные эвристики. Кроме того, при такой постановке отдельной проблемой является способ увязывания репутации с ожидаемым качеством ответа.

В [22] рассматривается задача назначения заданий, а именно — выбора сервисного узла (service node) в задачах мобильного краудсенсинга, однако здесь рассматривается уже многокритериальная задача оптимизации по ряду качественных критериев, актуальных для систем краудсенсинга (покрытие, задержки, достоверность и пр.). При этом предполагается, что известен вклад каждого из участников в значение каждой из мер качества (которые предполагаются аддитивными). В статье [23] предлагается оптимизационная модель для задач мобильного краудсенсинга, учитывающая вопросы приватности исполнителей, то есть рассматривается поиск компромисса между тремя метриками — приватность, полезность и эффективность.

Ко второй группе относятся работы [8] и [25]. Поскольку в них предполагается, что изначально информации о предполагаемой эффективности пользователей нет, в обеих работах используются элементы обучения с подкреплением, в частности модель многоруких бандитов (multi-armed bandits) [61]. Многорукые бандиты — это класс оптимизационных задач, в которых действия агента связаны с изначально неизвестными наградами, которые необходимо выявить посредством (зашумленных) экспериментов. Общей же целью оптимизации является максимизация полезности за счет выбора различных действий. Это соответствует ситуации выбора исполнителя задания — действиями агента (платформы краудсорсинга) являются назначение задач различным участникам, а результат выполнения (в частности его полезность) зависит от неизвестных заранее характеристик участника. Особенностью применения модели многоруких бандитов при распределении задач в системах краудсорсинга является то, что большинство классических вариантов этой модели (например, [62]) не рассматривают стоимость, которая может быть связана с каждым действием агента, а также возможные ограничения на количество заданий, которые могут быть назначены одному участнику (загрузка участника). Связи с этим в статьях, рассматривающих применение этой модели к распределению заданий в краудсорсинге, предлагаются собственные модификации.

В статье [8] рассматривается ситуация мобильного, пространственного краудсорсинга. Решается двоякая задача распределения заданий (с учетом ограниченного бюджета) и выявления

качества участников с использованием предлагаемого подхода, основанного на динамическом обучении (online learning). Эффективность демонстрируется с помощью имитационных моделей.

В работе [25] рассматривается ситуация выполнения работы с помощью участников, начальная эффективность которых неизвестна. Авторы подчеркивают, что ориентируются, в первую очередь, на сложные задания (краудсорсинг с помощью экспертов), мотивируя это тем, что такие условия применения краудсорсинга обладают следующими особенностями: существенная зависимость от уровня умений (отсюда особенная важность выбора лучших), открытый и динамичный рынок (нет базы всех исполнителей), сильный разброс цен на выполнение задания. Предлагаемое решение работает следующим образом: бюджет задачи разделяется на две части (в соответствии с задаваемым параметром  $\epsilon$ ), а исполнение разбито на две фазы. На первой фазе происходит выяснение эффективности участников (с использованием первой части бюджета), а на второй фазе задачи назначаются только наилучшим участникам (с помощью решения задачи о рюкзаке).

Важной характеристикой методов централизованного распределения является перечень факторов, учитываемых при распределении: ограничения бюджета, ограничения на загрузку исполнителей, временные показатели, а также набор критериев: качество, время и другое.

Таким образом:

- с прикладной точки зрения централизованное распределение заданий между участниками, осуществляемое посредством решения оптимизационной задачи, представляет собой удобную и расширяемую платформу для постановки и достижения различных целей (по качеству, оперативности, стоимости и пр.);

- решаемая при этом оптимизационная задача представляет собой вариант задачи о рюкзаке и (или множественной задачи о рюкзаке), является вычислительно трудной, что вынуждает использовать различные эвристики для ее решения. Эвристический подход может быть основан как на применении метаэвристик (генетическое программирование, имитация отжига и многих других), так и на создании специальных эвристик и в этом видится существенное поле для исследований;

- перспективным представляется использование методов обучения с подкреплением и активного обучения (в частности модель многоруких бандитов) для выяснения качеств участника, используемых в дальнейшем при формировании оптимального назначения.

**4.4. Теоретико-игровые методы.** Теория игр дает понятийный и математический аппарат, который, как показывают современные работы, может быть успешно использован для проектирования систем человеко-машинных вычислений. Использование теоретико-игровых методов позволяет рассматривать проблему обеспечения качества еще шире. Если методы, основанные на поиске согласия, предлагают только способ вывода истинных значений из накопленного набора данных, методы, связанные с проектированием потока работ, предлагают способы формирования потока работ, оптимального по ряду параметров, на основе известных (или оцениваемых в процессе) характеристик точности и прилежности исполнителей, то применение теоретико-игровых методов включает человека-исполнителя в рамки рассматриваемой системы, предлагая способ увязывания целей исполнителя (как активного элемента, действующего в своих интересах) с целями инициатора процесса вычислений (преследующего свои цели). И здесь проблема обеспечения качества находится в тесной связи с используемым механизмом стимулирования (*incentivization*), назначением которого как раз и является такое увязывание. В общем случае перечень задач, стоящих при разработке стратегии стимулирования, оказывается шире, чем обеспечение качества. Сюда входит, в первую очередь, привлечение участников к функционированию в системе, в том числе долговременному (если это предусмотрено — StackOverflow, Wikipedia и пр.).

Идея применения теоретико-игровых методов заключается в том, что (потенциальные) участники системы рассматриваются как рациональные агенты, основной целью которых является максимизация полезности от участия в системе. Функция полезности для участника определяется как разница между размером вознаграждения и стоимостью выполнения задания. Связь с качеством здесь в том, что модель может предусматривать разные стратегии участника, соответствующие различным уровням усилий, которые он может прикладывать, а от уровня усилий зависит качество получающихся результатов. Таким образом, участник осуществляет выбор уровня усилий (учитывая, что в системе есть и другие участники, действующие в своих интересах), чтобы максимизировать полезность. Задачей же разработчика системы является создание такого механизма распределения вознаграждения, чтобы приложение участниками максимальных усилий было равновесной по Нэшу стратегией.

В работе [63] таксономия методов стимулирования разбита на две части в зависимости от цели соответствующего механизма: предметно-ориентированные механизмы и общие (многоцелевые) механизмы.

Немного особняком стоит применение теоретико-игровых методов в краудсенсинге, поскольку участие в такого рода системах связано с расходованием определенных ресурсов — заряда батареи, сетевого трафика.

Применение теоретико-игровых методов связано и с определенными ограничениями. Например, необходимо считать известными функции зависимости качества от уровня усилий для участников, субъективную полезность для участников тех или иных видов вознаграждения, а также тот факт, что участники не всегда поступают рационально. Собственно, примерно по тем же причинам, по которым наряду с математической теорией игр появилась поведенческая теория игр [64]. Однако, несмотря на это, теоретико-игровые модели являются хорошим базисом для изучения и проектирования систем человеко-машинных вычислений.

В частности, теория игр и теоретико-игровые модели позволяют решать две важные задачи: во-первых, анализ существующих механизмов, лежащих в основе систем крауд-вычислений; во-вторых, разработка новых механизмов, удовлетворяющих требуемым свойствам. В список отобранных публикаций попали работы, посвященные решению каждой из перечисленных задач. Важные свойства, которым должны удовлетворять «хорошие» механизмы:

- Правдивость (Truthfulness). Ни один участник не может непосредственно увеличить свою полезность, делая ставку, отличающуюся от его стоимости выполнения задания.

- Индивидуальная рациональность (Individual rationality). Неотрицательность полезности для каждого участника.

- Вычислительная эффективность (Computational efficiency). Возможность вычислить результат (размер вознаграждения) за полиномиальное время. В [31] вычислительная эффективность также трактуется как возможность работы в реальном времени.

- Сбалансированность бюджета (Budget-balance) или прибыльность для организатора (Platform profitable). Неотрицательность полезности для организатора (в данном случае, платформы крауд-вычислений).

Одной из первых публикаций по применению теории игр для анализа устройства систем крауд-вычислений является [28], в которой предлагается обобщенная аналитическая модель игры ESP и производится анализ параметров этой аналитической модели на качество получаемых результатов. Игра ESP предложена в [65] и предназначена для аннотирования изображений. Суть игры в том, что двум участникам предъявляется одно и то же изображение и каждый из них должен

описать это изображение с помощью ключевых слов. Количество набранных участниками очков зависит от того, совпали ли ключевые слова. Таким образом, каждый из участников, преследуя цель набрать наибольшее количество очков, пытается выбрать те слова, которые, как он считает, вероятнее всего выберет другой участник.

В статье [27] производится анализ распространенных в области краудсорсинга микрозадач механизмов (случайного назначения, дополненного мажоритарным голосованием, и централизованного назначения с последующей проверкой) с позиций теории игр и показываются их недостатки. Предлагаемый же авторами механизм основан на том, что участникам назначаются тренировочные задания, способные повышать качество их результатов за счет эффекта обучения. Участник моделируется вероятностью того, что результаты его решений будут приемлемого качества, предполагается также, что множество решений бесконечно и вероятность того, что два участника получают одинаковое ложное решение, равна 0. Стратегии участника связаны с уровнем усилий, которые он может приложить, известна эта функция, удовлетворяющая ряду условий. Выигрыш (полезность) участника при наличии механизма  $M$  зависит от выбранного качества решения и качества решений, полученных от других участников. Механизмы сравниваются и оцениваются по их симметрическим равновесным по Нэшу стратегиям (symmetric Nash equilibria), среди которых наибольший интерес представляют те, в которых участники имеют стимул для того, чтобы генерировать ответы наилучшего возможного качества, прилагая максимум усилий. С другой стороны, механизм также оценивается по стоимости для организатора (рекрутера). В частности, по способности гарантированно «укладываться» в любой неотрицательный заранее известный бюджет (подробнее про эту постановку можно почитать в [27]).

В работе [29] предлагается механизм стимулирования для мобильного краудсенсинга, основанный на аукционе по качеству. Основной областью применения являются такие приложения краудсенсинга, в которых вознаграждение участника зависит от качества полученных от него данных, а не от времени работы и других параметров. В работе теоретически доказан ряд свойств предложенного механизма: правдивость, индивидуальная рациональность, прибыльность для организатора. Показано также применение предложенного механизма при разработке приложения для позиционирования внутри помещений, основанного на методе радиоотпечатков Wi-Fi.

В статье [30] предлагается механизм стимулирования для мобильного краудсенсинга. В отличие от множества других механизмов, ориентирующихся, в первую очередь, на измерение

качества вклада определенного пользователя и подборе адекватного стимула для него, в данной статье предлагается механизм социального стимулирования, направленный на стимулирование социального окружения участника, выполняющего задание.

Отдельную и относительно хорошо изученную группу механизмов составляют аукционы.

В статье [32] предлагается механизм стимулирования для участников систем краудсенсинга, комбинирующий элементы обратных аукционов (reverse auctions) и аукционов Викри (Vikrey auctions). Показано, что механизм является вычислительно эффективным, индивидуально рациональным, сбалансированным по бюджету и правдивым. Имитационные эксперименты показали, что применение механизма позволяет улучшить качество данных, получаемых от системы. Основной сценарий, рассматриваемый в статье, заключается в том, что участники могут делать ставки на выполнение заданий (указывая, за какую компенсацию они согласны выполнить задание), по результатам аукциона задание назначается участникам, а факт получения вознаграждения зависит от качества данных, предоставленных участником. При этом основной фокус статьи сосредоточен на том, чтобы разработанный механизм не допускал злонамеренного искажения ставок.

В работе [26] предложен концептуальный каркас (framework) стимулирования для краудсенсинга, интегрирующий три популярных метода стимулирования: обратные аукционы, геймификацию и систему репутации, и затрагивающая все три стороны, заинтересованные в процессе краудсенсинга: участников, рекрутеров и саму платформу. Особенностью этой работы является включение рекрутера в общую схему. В частности, решается проблема правдивости «приема» заданий рекрутером, для чего задача по приему заданий назначается участникам, используя геймификацию в качестве механизма, и в зависимости от решения участников рекрутер либо получает, либо не получает возмещение.

В статье [31] рассматривается задача создания механизма, удовлетворяющего шести необходимым свойствам: вычислительная эффективность, индивидуальная рациональность, прибыльность для организатора, правдивость (участники сообщают истинную стоимость решения задачи и истинные сроки ее выполнения), возможность каждого участника выиграть аукцион, близка к оптимальному, при полностью известной информации об участниках. Общая модель взаимодействия такая — участник характеризуется временем прихода, временем ухода, стоимостью выполнения задания и набором заданий,

которые он может выполнить (это составляет тип участника). Участник делает ставку, и среда должна решить, покупает ли она услуги участника за эту ставку до момента ухода участника. При этом рекрутер преследует цель получения максимальной полезности в рамках своих денежных ограничений. Однако качество здесь сводится только лишь к требуемому количеству повторений задания.

Таким образом:

– теория игр и в особенности ее раздел, посвященный дизайну механизмов (mechanism design), дают очень хорошую концептуальную базу, позволяющую производить анализ систем крауд-вычислений, а также проектировать системы, рациональные участники которых заинтересованы в качественном выполнении заданий. И упоминание рациональности здесь очень важно, поскольку при ее отсутствии и/или при целенаправленных попытках нарушения работы системы теоретико-игровые модели перестают адекватно описывать поведение участников. Вместе с тем качество определяется не только заинтересованностью участника в выполнении задания, поэтому теоретико-игровые методы, по всей видимости, должны сочетаться с методами других групп;

– к перспективным направлениям развития этой группы методов относится поведенческая теория игр, объединение теоретико-игровых моделей систем крауд-вычислений с практическими исследованиями демографии площадок крауд-вычислений, качественными и количественными моделями поведения участников таких систем.

#### **4.5. Методы, основанные на учете свойств заданий.**

Исходными данными для большинства методов и подходов к обеспечению качества является информация об участниках и множество накопленных результатов выполнения заданий. Характеристики самих заданий практически не учитываются, и это объяснимо, поскольку крауд-вычисления традиционно применяются в тех областях, где машинная обработка информации оказывается недостаточно надежной. Однако среди отобранных источников оказалось и несколько статей, учитывающих свойства заданий. Эти статьи позволяют наметить область применимости подобного подхода.

В статье [34] краудсорсинг используется для разметки научных данных (астрономических снимков). Авторы предлагают оригинальную модель, в которой предлагается оценивать непротиворечивость результатов, полученных от самого участника, на основе того факта, что классификаторы хорошо работают с незашумленными данными. Для этого авторы используют распространенную модель автоматической классификации астрономических объектов по визуальным признакам и

обучают ее на качественных данных, чтобы оценить границы точности предсказаний. Затем модель обучается на подмножестве данных, размеченных участником, и проверяется ее предиктивная «сила». Если оказывается, что она существенно меньше, чем предиктивная «сила» модели, обученной на реальных данных, то делается вывод, что разметка, полученная от пользователя, является зашумленной. Ограничение метода в том, что для тренировки модели требуется большое количество аннотаций (хорошие показатели у авторов получаются, например, при 100 аннотаций).

Есть также попытки адаптации методов очистки данных к идентификации противоречивых результатов исполнения заданий участниками системы крауд-вычислений. Так, в статье [33] предлагается несколько методов очистки — основанный на известном методе «полировки» [66], основанный на самообучающихся классификаторах и основанный на кластерном анализе. Интересно, что предлагаемая авторами разновидность метода «полировки» и очистка, основанная на самообучающихся классификаторах, по принципу работы напоминают различные методы согласования, с тем отличием, что источниками меток являются не люди, а модели классификации, обученные на разных подмножествах. Очистка же с помощью кластерного анализа требует, чтобы в рамках одного кластера объектов (заданий) результаты, полученные от людей, были одинаковы. Схожие подходы рассмотрены и проанализированы в [1].

Обобщая эти работы можно отметить, что свойства самих заданий (или результат машинной обработки информации, связанной с заданиями) актуально использовать по крайней мере в двух случаях:

1) Если для них существуют качественные математические модели, позволяющие получить точность обработки, сопоставимую с точностью обработки человеком.

2) Когда часть заданий (вызывающую наибольшие противоречия) можно не учитывать в дальнейшей обработке (например, при подготовке данных для машинного обучения устранение лишних данных — фильтрация — ведет часто лишь к большей устойчивости моделей).

**4.6. Методы, основанные на анализе действий пользователя и воздействии на него.** Отдельное направление методов обеспечения качества связано с анализом действий пользователя в ходе выполнения заданий и влиянием на процесс выполнения заданий.

Основной идеей здесь является построение классификатора, позволяющего связать поведение пользователя (движение указателя, активизация тех или иных управляющих элементов формы, временные задержки, сложность ввода, следование «мягким» правилам при

выполнении задания) с его «добросовестностью» при выполнении заданий. Для обучения такой модели, как правило, используются задания с заранее известными результатами. Сами признаки, определяющие вид модели классификации, в значительной мере определяются видом заданий. Примером такой работы является [36].

В работе [35] для обеспечения качества применяется подход, связанный с мониторингом психических характеристик участника (усталости) с помощью эталонных заданий. Авторы внедряют задания с известным результатом для того, чтобы обнаружить момент, когда из-за усталости качество результатов участника снижается, и прекращают назначать ему задания.

Помимо работ, в которых происходит мониторинг действий пользователя, в эту же группу можно отнести и работы, в которых подчеркиваются различные аспекты проектирования заданий, которые могут повлиять на качество получаемых результатов.

Статья [37] посвящена оригинальному, скорее, организационному способу повышения качества результатов в краудсорсинге микрозаданий — применению техники прайминга (психологический и маркетинговый термин, обозначающий фиксирование установки [67]). На практике это проявлялось в том, что заданиям сопутствовали изображения и аудио-фрагменты, призванные вызвать определенное эмоциональное состояние исполнителя (сформировать установку на качественное выполнение работы). Приведенные в статье результаты показывают, что использование техник прайминга действительно позволяет добиться краткосрочного улучшения результатов.

В рассмотренной уже ранее статье [21] вспомогательным методом повышения качества являлось помещение на форму выполнения задания ссылок, поясняющих специальные термины. И хотя это решение представляется разумным, в статье нет экспериментальных свидетельств о влиянии решения на качество результатов.

**5. Заключение.** Проведен аналитический обзор современных (вышедших после 2012 года) статей библиографических баз ScienceDirect и IEEE Xplore, посвященных методам обеспечения качества результатов при использовании крауд-вычислений. Выявлены основные направления, по которым происходит развитие таких методов. Показано, что наибольшее внимание научного сообщества сосредоточено в двух во многом противоположных направлениях: обеспечение качества, основанное на согласовании данных, полученных от разных участников, и обеспечение качества за счет использования теоретико-игровых методов. Противоположность же этих направлений связана с

тем, что методы, основанные на согласовании, рассматривают задачу в максимально узкой постановке — как наиболее правильно интерпретировать собранные данные, а теоретико-игровые методы, напротив, предлагают концептуальную платформу для проектирования и разработки таких систем.

Выявленные подходы сопоставлены с основными этапами процесса применения крауд-вычислений к решению задачи (таблица 2). Понимание места того или иного подхода в рамках процесса реализации крауд-вычислений имеет существенное значение, потому что различные платформы, с помощью которых реализуются подобные системы, обладают разными особенностями работы и возможностями, что может накладывать ограничения на допустимые методы. Например, популярная платформа краудсорсинга микрозаданий Amazon Mechanical Turk осуществляет распределение заданий в так называемом «пулл»-режиме, то есть участники сами выбирают задания. Следовательно, методы, основанные на централизованном распределении оказываются в этом случае неприменимы. Наименьшие ограничения в этом смысле у методов согласования, поскольку для их применения достаточно лишь набора данных, полученного от участников.

В работах [68, 69] делается попытка выделения причин получения некачественных результатов в крауд-вычислениях. Среди основных факторов, влияющих на качество, названы компетентность, заинтересованность в выполнении заданий (мотивация), ясность представления заданий, наличие или отсутствие «злого умысла» (целенаправленного подрыва работы системы). Различные классы методов в основном учитывают воздействие лишь какого-то подмножества факторов. Так, учет мотивационных факторов происходит с помощью теоретико-игровых методов, выявление и использование сведений о компетентности участников происходит в большинстве методов разных групп, наконец, свойства задания частично рассматриваются в последней группе. Таким образом, комплексный подход к проблеме обеспечения качества должен базироваться на сочетании различных методов.

Следует также отметить, что в настоящий момент отсутствуют публикации, в которых сравнивались бы принципиально различные подходы к обеспечению качества результатов, получаемых с помощью крауд-вычислений. В то же время только такие комплексные работы способны заложить основы рационального и научно обоснованного подхода к принятию проектных решений при создании человеко-машинных систем, использующих данную технологию.

## Литература

1. *Li C. et al.* Noise filtering to improve data and model quality for crowdsourcing // Knowledge-Based Syst. 2016. vol. 107. pp. 96–103.
2. *Bernstein A., Klein M., Malone T.W.* Programming the global brain // Commun. ACM. 2012. vol. 55. no. 5. pp. 41–43.
3. *Kazai G., Kamps J., Milic-Frayling N.* Worker types and personality traits in crowdsourcing relevance labels // Proc. 20th ACM Int. Conf. Inf. Knowl. Manag. (CIKM '11). USA: ACM Press. 2011. pp. 1941–1944.
4. *Kitchenham B.* Procedures for Performing Systematic Reviews // Keele. 2004. vol. 33. no. 2004. pp. 1–26.
5. *Kitchenham B., Charters S.* Guidelines for performing Systematic Literature Reviews in Software Engineering. Technical Report EBSE-2007-01. 2007. 65 p.
6. *Zhang J., Sheng V, Li Q., Wu J., Wu X.* Consensus algorithms for biased labeling in crowdsourcing // Information Sciences. 2017. vol. 382. pp. 254–273.
7. *Matsunaga A., Mast A., Fortes J.A.B.* Workforce-efficient consensus in crowdsourced transcription of biocollections information // Future Generation Computer Systems. 2016. vol. 56. pp. 526–536.
8. *Han K., Zhang C., Luo J.* Taming the Uncertainty: Budget Limited Robust Crowdsensing Through Online Learning // Biological Cybernetics. 2016. vol. 24. no. 3. pp. 1462–1475.
9. *Dang D., Liu Y., Zhang X., Huang S.* Crowdsourcing Worker Quality Evaluation Algorithm on MapReduce for Big Data Applications // IEEE Transactions on Parallel and Distributed Systems. 2016. vol. 27. no. 7. pp. 1879–1888.
10. *Duan L., Oyama S., Sato H., Kurihara M.* Separate or joint? Estimation of multiple labels from crowdsourced annotations // Expert Systems with Applications. 2014. vol. 41. no. 13. pp. 5723–5732.
11. *Kara Y.E., Genc G., Aran O., Akarun L.* Modeling annotator behaviors for crowd labeling // Neurocomputing. 2015. vol. 160. pp. 141–156.
12. *Kubota T., Aritsugi M.* Assignment strategies for ground truths in the crowdsourcing of labeling tasks // The Journal of Systems and Software. 2017. vol. 126. pp. 113–126.
13. *Otani N., Baba Y., Kashima H.* Quality control of crowdsourced classification using hierarchical class structures // Expert Systems with Applications. 2016. vol. 58. pp. 155–163.
14. *Vempaty A., Varshney L.R., Varshney P.K.* Reliable Crowdsourcing for Multi-Class Labeling // IEEE Journal of Selected Topics in Signal Process. 2014. vol. 8. no. 4. pp. 667–679.
15. *Vuurens J.B.P., de Vries A.P.* Obtaining High-Quality Relevance Judgments Using Crowdsourcing // IEEE Internet Comput. 2012. vol. 16. no. 5. pp. 20–27.
16. *Zhang J., Sheng V., Wu J., Wu X.* Multi-Class Ground Truth Inference in Crowdsourcing with Clustering // IEEE Transactions on Knowledge and Data Engineering. 2016. vol. 28. no. 4. pp. 1080–1085.
17. *Zhang J., Wu X., Sheng V.S.* Imbalanced Multiple Noisy Labeling // IEEE Transactions on Knowledge and Data Engineering. 2015. vol. 27. no. 2. pp. 489–503.
18. *Dai P., Lin C., Weld D.* POMDP-based control of workflows for crowdsourcing // Artificial Intelligence. 2013. vol. 202. pp. 52–85.
19. *Hirth M., Hofffeld T., Tran-Gia P.* Analyzing costs and accuracy of validation mechanisms for crowdsourcing platforms // Mathematical and Computer Modeling. 2013. vol. 57. no. 11–12. pp. 2918–2932.
20. *Lee J., Lee D., Hwang S.* CrowdK: Answering top- k queries with crowdsourcing // Information Sciences. 2017. vol. 399. pp. 98–120.
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. no. C. pp. 20–34.

22. *Ding S., He X., Wang J.* Multiobjective Optimization Model for Service Node Selection Based on a Tradeoff Between Quality of Service and Resource Consumption in Mobile Crowd Sensing // IEEE Internet of Things Journal. 2017. vol. 4. no. 1. pp. 258–268.
23. *Gong Y. et al.* Optimal Task Recommendation for Mobile Crowdsourcing With Privacy Control // IEEE Internet of Things Journal. 2016. vol. 3. no. 5. pp. 745–756.
24. *Miao C., Yu H., Shen Z., Leung C.* Balancing quality and budget considerations in mobile crowdsourcing // Decision Support Systems. 2016. vol. 90. pp. 56–64.
25. *Tran-Thanh L., Stein S., Rogers A., Jennings N.* Efficient crowdsourcing of unknown experts using bounded multi-armed bandits // Artificial Intelligence. 2014. vol. 214. pp. 89–111.
26. *Dai W., Wang Y., Jin Q., Ma J.* An Integrated Incentive Framework for Mobile Crowdsourced Sensing // Tsinghua Science and Technology. 2016. vol. 21. no. 2. pp. 146–156.
27. *Gao Y., Chen Y., Liu K.J.R.* On Cost-Effective Incentive Mechanisms in Microtask Crowdsourcing // IEEE Transactions on Computational Intelligence and AI in Games. 2015. vol. 7. no. 1. pp. 3–15.
28. *Wang B.C., Lin C.W., Chen K.T., Chen L.J.* An analytical model for generalized ESP games // Knowledge-Based Systems. 2012. vol. 34. pp. 114–127.
29. *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.
30. *Yang G., He S., Shi Z., Chen J.* Promoting Cooperation by the Social Incentive Mechanism in Mobile Crowdsensing // IEEE Communications Magazine. 2017. vol. 55. no. 3. pp. 86–92.
31. *Zhao D., Li X., Ma H.* Budget-Feasible Online Incentive Mechanisms for Crowdsourcing Tasks Truthfully // IEEE/ACM Transactions on Networking. 2016. vol. 24. no. 2. pp. 647–661.
32. *Zhu X. et al.* A Fair Incentive Mechanism for Crowdsourcing in Crowd Sensing // IEEE Internet of Things Journal. 2016. vol. 3. no. 6. pp. 1364–1372.
33. *Nicholson B., Sheng V.S., Zhang J.* Label noise correction and application in crowdsourcing // Expert Systems with Applications. 2016. vol. 66. pp. 149–162.
34. *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.
35. *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.
36. *Mok R.K.P., Chang R.K.C., Li W.* Detecting Low-Quality Workers in QoE Crowdtesting: A Worker Behavior-Based Approach // IEEE Transactions on Multimedia. 2017. vol. 19. no. 3. pp. 530–543.
37. *Morris R.R., Dontcheva M., Gerber E.M.* Priming for Better Performance in Microtask Crowdsourcing Environments // IEEE Internet Computing. 2012. vol. 16. no. 5. pp. 13–19.
38. *Allahbakhsh M. et al.* Quality Control in Crowdsourcing Systems: Issues and Directions // IEEE Internet Computing. 2013. vol. 17. no. 2. pp. 76–81.
39. *Chittilappilly A.I., Chen L., Amer-Yahia S.* A Survey of General-Purpose Crowdsourcing Techniques // IEEE Transactions on Knowledge and Data Engineering. 2016. vol. 28. no. 9. pp. 2246–2266.
40. *Li G., Wang J., Zheng Y., Franklin M.J.* Crowdsourced Data Management: A Survey // IEEE Transactions on Knowledge and Data Engineering. 2016. vol. 28. no. 9. pp. 2296–2319.
41. *Liu C.H. et al.* QoI-Aware Energy-Efficient Participatory Crowdsourcing // IEEE Sensors Journal. 2013. vol. 13. no. 10. pp. 3742–3753.
42. *Liu C.H. et al.* Toward QoI and Energy Efficiency in Participatory Crowdsourcing // IEEE Transactions on Vehicular Technology. 2015. vol. 64. no. 10. pp. 4684–4700.

43. *Yung D., Li M.-L., Chang S.* Evolutionary approach for crowdsourcing quality control // *Journal of Visual Languages and Computing*. 2014. vol. 25. no. 6. pp. 879–890.
44. *Zhang B. et al.* Privacy-preserving QoI-aware participant coordination for mobile crowdsourcing // *Computing Networks*. 2016. vol. 101. pp. 29–41.
45. *Zhang X. et al.* Robust Trajectory Estimation for Crowdsourcing-Based Mobile Applications // *IEEE Transactions on Parallel and Distributed Systems*. 2014. vol. 25. no. 7. pp. 1876–1885.
46. *Sheng V.S., Provost F., Ipeirotis P.G.* Get another label? improving data quality and data mining using multiple, noisy labelers // *Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD 08)*. 2008. pp. 614–622.
47. *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.
48. *Raykar V.C. et al.* Learning from crowds // *Journal of Machine Learning Research*. 2010. vol. 11. pp. 1297–1322.
49. *Karger D.R., Oh S., Shah D.* Iterative learning for reliable crowdsourcing systems // *Proceedings of the 24th International Conference on Neural Information Processing Systems*. 2011. pp. 1953–1961.
50. *Zhang Y., Chen X., Zhou D., Jordan M.I.* Spectral methods meet EM: A provably optimal algorithm for crowdsourcing // *Proceedings of the 27th International Conference on Neural Information Processing Systems, NIPS'14, Advances in Neural Information Processing Systems*. 2014. vol. 2. pp. 1260–1268.
51. *Demartini G., Dfallah D.E., Cudré-Mauroux P.* ZenCrowd: leveraging probabilistic reasoning and crowdsourcing techniques for large-scale entity linking // *Proceedings of the 21st international conference on World Wide Web (WWW '12)*. 2012. pp. 469–478.
52. *Gupta M.R., Chen Y.* Theory and Use of the EM Algorithm // *Foundations and Trends in Signal Processing*. 2010. vol. 4. no. 3. pp. 223–296.
53. *McLachlan G.J., Krishnan T.* *The EM algorithm and extensions*: 2nd ed. // *Wiley-Interscience*, 2008. 359 p.
54. *Sheshadri A., Lease M.* Square: A benchmark for research on computing crowd consensus // *Proceedings of the 1st AAAI Conference on Human Computations and Crowdsourcing*. 2013. pp. 156–164.
55. *Quoc Viet Hung N., Tam N.T., Tran L.N., Aberer K.* An Evaluation of Aggregation Techniques in Crowdsourcing // *Web Information Systems Engineering – WISE 2013*. 2013. LNCS 8181. pp. 1–15.
56. *Khosravifar B., Bentahar J., Gomrokchi M., Alam R.* CRM: An efficient trust and reputation model for agent computing // *Knowledge-Based Systems*. 2012. vol. 30. pp. 1–16.
57. *Fang H., Guo G., Zhang J.* Multi-faceted trust and distrust prediction for recommender systems // *Decision Support Systems*. 2015. vol. 71. pp. 37–47.
58. *Wahab O.A., Bentahar J., Otrok H., Mourad A.* A survey on trust and reputation models for Web services: Single, composite, and communities // *Decision Support Systems*. 2015. vol. 74. pp. 121–134.
59. *Yu H. et al.* Filtering trust opinions through reinforcement learning // *Decision Support Systems*. 2014. vol. 66. p. 102–113.
60. *Liu S., Yu H., Miao C., Kot A.C.* A fuzzy logic based reputation model against unfair ratings // *Proceedings of the 2013 international conference on Autonomous agents and multi-agent systems*. 2013. pp. 821–828.
61. *Auer P., Cesa-Bianchi N., Fischer P.* Finite-time Analysis of the Multiarmed Bandit Problem // *Machine Learning*. 2002. vol. 47. no. 2/3. pp. 235–256.
62. *Bubeck S., Cesa-Bianchi N.* Regret analysis of stochastic and nonstochastic multi-armed bandit problems. 2012. 138 p.

63. *Restuccia F., Das S.K., Payton J.* Incentive Mechanisms for Participatory Sensing: Survey and Research Challenges // *ACM Transactions on Sensor Networks*. 2016. vol. 12. no. 2. Article no. 13.
64. *Camerer C.F.* Behavioral game theory: Predicting human behavior in strategic situations // *Advances in Behavioral Economics*. 2011. pp. 374–392.
65. *Von Ahn L., Dabbish L.* Labeling images with a computer game // *Proceedings of Conference on Human Factors in Computing Systems*. 2004. pp. 319–326.
66. *Teng C.M.* Correcting noisy data // *Proceedings of the Sixteenth International Conference on Machine Learning*. 1999. pp. 239–248.
67. *Фаликман М.В., Койфман А.Я.* Виды прайминга в исследованиях восприятия и перцептивного внимания. Часть 1 // *Вестник Московского университета. Серия 14. Психология*. 2005. № 3. С. 86–97.
68. *Frenay B., Verleysen M.* Classification in the Presence of Label Noise: A Survey // *IEEE Transactions on Neural Networks and Learning Systems*. 2014. vol. 25. no. 5. pp. 845–869.
69. *Eickhoff C., de Vries A.P.* Increasing cheat robustness of crowdsourcing tasks // *Information Retrieval*. 2013. vol. 16. no. 2. pp. 121–137.

**Пономарев Андрей Васильевич** — к-т техн. наук, старший научный сотрудник лаборатории интегрированных систем автоматизации, Федеральное государственное бюджетное учреждение науки Санкт-Петербургского института информатики и автоматизации Российской академии наук (СПИИРАН). Область научных интересов: системы коллективного интеллекта, социальные вычисления, краудсорсинг, рекомендующие системы, технологии поддержки принятия решений, дискретная оптимизация. Число научных публикаций — 46. [ponomarev@iias.spb.su](mailto:ponomarev@iias.spb.su); 14-я линия В.О., 39, Санкт-Петербург, 199178; р.т.: +7(812)328-8071, Факс: +7(812)328-4450.

**Поддержка исследований.** Работа выполнена при финансовой поддержке РФФИ (проекты № 16-37-60107 и № 16-07-00466).

A.V. PONOMAREV  
**QUALITY CONTROL METHODS IN CROWD COMPUTING:  
 LITERATURE REVIEW**

---

*Ponomarev A.V. Quality Control Methods in Crowd Computing: Literature Review.*

**Abstract.** Currently, crowd computing is gaining popularity. However, the quality of results obtained by means of crowd computing is often unpredictable, and that fact puts limits on the practical applicability of this technology. Therefore, systematization of information about modern methods for quality control in crowd computing is an important task that can pave the way to new research efforts in this area and therefore widen the scope of its applicability. The paper discusses the results of systematic literature review of journal articles from ScienceDirect and IEEE Xplore bibliographic databases, published after 2012. The paper also identifies main directions in crowd computing quality control, corresponding models and assumptions. In particular, it shows that most scientific attention is concentrated around consensus methods and game theoretic methods.

**Keywords:** crowd computing, human-machine computing, crowdsourcing, social computing, human-machine systems, human factors, systematic literature review.

---

**Ponomarev Andrew Vasil'evich** — Ph.D., senior researcher of computer aided integrated systems laboratory, St.Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences (SPIIRAS). Research interests: collective intelligence systems, social computing, crowdsourcing, machine learning, recommender systems, decision support technologies, discrete optimization. The number of publications — 46. [ponomarev@iias.spb.su](mailto:ponomarev@iias.spb.su); 39, 14-th Line V.O., St. Petersburg, 199178, Russia; office phone: +7(812)328-8071, Fax: +7(812)328-4450.

**Acknowledgements.** The research is financially supported by RFBR (projects no. 16-37-60107 and 16-07-00466).

### References

1. Li C. et al. Noise filtering to improve data and model quality for crowdsourcing // *Knowledge-Based Syst.* 2016. vol. 107. pp. 96–103.
2. Bernstein A., Klein M., Malone T.W. Programming the global brain // *Commun. ACM.* 2012. vol. 55. no. 5. pp. 41–43.
3. Kazai G., Kamps J., Milic-Frayling N. Worker types and personality traits in crowdsourcing relevance labels // Proc. 20th ACM Int. Conf. Inf. Knowl. Manag. (CIKM '11). USA: ACM Press. 2011. pp. 1941–1944.
4. Kitchenham B. Procedures for Performing Systematic Reviews // Keele. 2004. vol. 33. no. 2004. pp. 1–26.
5. Kitchenham B., Charters S. Guidelines for performing Systematic Literature Reviews in Software Engineering. Technical Report EBSE-2007-01. 2007. 65 p.
6. Zhang J. et al. Consensus algorithms for biased labeling in crowdsourcing. *Information Sciences.* 2017. vol. 382. pp. 254–273.
7. Matsunaga A., Mast A., Fortes J.A.B. Workforce-efficient consensus in crowdsourced transcription of biocollections information. *Future Generation Computer Systems.* 2016. vol. 56. pp. 526–536.
8. Han K., Zhang C., Luo J. Taming the Uncertainty: Budget Limited Robust Crowdsensing Through Online Learning. *Biological Cybernetics.* 2016. vol. 24. no. 3.

- pp. 1462–1475.
9. Dang D., Liu Y., Zhang X., Huang S. Crowdsourcing Worker Quality Evaluation Algorithm on MapReduce for Big Data Applications. *IEEE Transactions on Parallel and Distributed Systems*. 2016. vol. 27. no. 7. pp. 1879–1888.
  10. Duan L., Oyama S., Sato H., Kurihara M. Separate or joint? Estimation of multiple labels from crowdsourced annotations. *Expert Systems with Applications*. 2014. vol. 41. no. 13. pp. 5723–5732.
  11. Kara Y.E., Genc G., Aran O., Akarun L. Modeling annotator behaviors for crowd labeling. *Neurocomputing*. 2015. vol. 160. pp. 141–156.
  12. Kubota T., Aritsugi M. Assignment strategies for ground truths in the crowdsourcing of labeling tasks. *The Journal of Systems and Software*. 2017. vol. 126. pp. 113–126.
  13. Otani N., Baba Y., Kashima H. Quality control of crowdsourced classification using hierarchical class structures. *Expert Systems with Applications*. 2016. vol. 58. pp. 155–163.
  14. Vempaty A., Varshney L.R., Varshney P.K. Reliable Crowdsourcing for Multi-Class Labeling. *IEEE Journal of Selected Topics in Signal Process*. 2014. vol. 8. no. 4. pp. 667–679.
  15. Vuurens J.B.P., de Vries A.P. Obtaining High-Quality Relevance Judgments Using Crowdsourcing. *IEEE Internet Computing*. 2012. vol. 16. No. 5. pp. 20–27.
  16. Zhang J., Sheng V., Wu J., Wu X. Multi-Class Ground Truth Inference in Crowdsourcing with Clustering. *IEEE Transactions on Knowledge and Data Engineering*. 2016. vol. 28. no. 4. pp. 1080–1085.
  17. Zhang J., Wu X., Sheng V.S. Imbalanced Multiple Noisy Labeling. *IEEE Transactions on Knowledge and Data Engineering*. 2015. vol. 27. no. 2. pp. 489–503.
  18. Dai P., Lin C., Weld D. POMDP-based control of workflows for crowdsourcing. *Artificial Intelligence*. 2013. vol. 202. pp. 52–85.
  19. Hirth M., Hoßfeld T., Tran-Gia P. Analyzing costs and accuracy of validation mechanisms for crowdsourcing platforms. *Mathematical and Computer Modeling*. 2013. vol. 57. no. 11–12. pp. 2918–2932.
  20. Lee J., Lee D., Hwang S. CrowdK: Answering top- k queries with crowdsourcing. *Information Sciences*. 2017. vol. 399. pp. 98–120.
  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. no. C. pp. 20–34.
  22. Ding S., He X., Wang J. Multiobjective Optimization Model for Service Node Selection Based on a Tradeoff Between Quality of Service and Resource Consumption in Mobile Crowd Sensing. *IEEE Internet of Things Journal*. 2017. vol. 4. no. 1. pp. 258–268.
  23. Gong Y. et al. Optimal Task Recommendation for Mobile Crowdsourcing With Privacy Control. *IEEE Internet of Things Journal*. 2016. vol. 3. no. 5. pp. 745–756.
  24. Miao C., Yu H., Shen Z., Leung C. Balancing quality and budget considerations in mobile crowdsourcing. *Decision Support Systems*. 2016. vol. 90. pp. 56–64.
  25. Tran-Thanh L., Stein S., Rogers A., Jennings N. Efficient crowdsourcing of unknown experts using bounded multi-armed bandits. *Artificial Intelligence*. 2014. vol. 214. pp. 89–111.
  26. Dai W., Wang Y., Jin Q., Ma J. An Integrated Incentive Framework for Mobile Crowdsourced Sensing. *Tsinghua Science and Technology*. 2016. vol. 21. no. 2. pp. 146–156.
  27. Gao Y., Chen Y., Liu K.J.R. On Cost-Effective Incentive Mechanisms in Microtask Crowdsourcing. *IEEE Transactions on Computational Intelligence and AI in Games*. 2015. vol. 7. no. 1. pp. 3–15.
  28. Wang B.C., Lin C.W., Chen K.T., Chen L.J. An analytical model for generalized ESP games. *Knowledge-Based Systems*. 2012. vol. 34. pp. 114–127.
  29. 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.
30. Yang G., He S., Shi Z., Chen J. Promoting Cooperation by the Social Incentive Mechanism in Mobile Crowdsensing. *IEEE Communications Magazine*. 2017. vol. 55. no. 3. pp. 86–92.
  31. Zhao D., Li X., Ma H. Budget-Feasible Online Incentive Mechanisms for Crowdsourcing Tasks Truthfully. *IEEE/ACM Transactions on Networking*. 2016. vol. 24. no. 2. pp. 647–661.
  32. Zhu X. et al. A Fair Incentive Mechanism for Crowdsourcing in Crowd Sensing. *IEEE Internet of Things Journal*. 2016. vol. 3. no. 6. pp. 1364–1372.
  33. Nicholson B., Sheng V.S., Zhang J. Label noise correction and application in crowdsourcing. *Expert Systems with Applications*. 2016. vol. 66. pp. 149–162.
  34. 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.
  35. 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.
  36. Mok R.K.P., Chang R.K.C., Li W. Detecting Low-Quality Workers in QoE Crowdtesting: A Worker Behavior-Based Approach. *IEEE Transactions on Multimedia*. 2017. vol. 19. no. 3. pp. 530–543.
  37. Morris R.R., Dontcheva M., Gerber E.M. Priming for Better Performance in Microtask Crowdsourcing Environments. *IEEE Internet Computing*. 2012. vol. 16. no. 5. pp. 13–19.
  38. Allahbakhsh M. et al. Quality Control in Crowdsourcing Systems: Issues and Directions. *IEEE Internet Computing*. 2013. vol. 17. no. 2. pp. 76–81.
  39. Chittilappilly A.I., Chen L., Amer-Yahia S. A Survey of General-Purpose Crowdsourcing Techniques. *IEEE Transactions on Knowledge and Data Engineering*. 2016. vol. 28. no. 9. pp. 2246–2266.
  40. Li G., Wang J., Zheng Y., Franklin M.J. Crowdsourced Data Management: A Survey. *IEEE Transactions on Knowledge and Data Engineering*. 2016. vol. 28. no. 9. pp. 2296–2319.
  41. Liu C.H. et al. QoI-Aware Energy-Efficient Participatory Crowdsourcing. *IEEE Sensors Journal*. 2013. vol. 13. no. 10. pp. 3742–3753.
  42. Liu C.H. et al. Toward QoI and Energy Efficiency in Participatory Crowdsourcing. *IEEE Transactions on Vehicular Technology*. 2015. vol. 64. no. 10. pp. 4684–4700.
  43. Yung D., Li M.-L., Chang S. Evolutionary approach for crowdsourcing quality control. *Journal of Visual Languages and Computing*. 2014. vol. 25. no. 6. pp. 879–890.
  44. Zhang B. et al. Privacy-preserving QoI-aware participant coordination for mobile crowdsourcing. *Computing Networks*. 2016. vol. 101. pp. 29–41.
  45. Zhang X. et al. Robust Trajectory Estimation for Crowdsourcing-Based Mobile Applications. *IEEE Transactions on Parallel and Distributed Systems*. 2014. vol. 25. no. 7. pp. 1876–1885.
  46. Sheng V.S., Provost F., Ipeirotis P.G. Get another label? improving data quality and data mining using multiple, noisy labelers. Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD 08). 2008. pp. 614–622.
  47. 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.
  48. Raykar V.C. et al. Learning from crowds. *Journal of Machine Learning Research*. 2010. vol. 11. pp. 1297–1322.
  49. Karger D.R., Oh S., Shah D. Iterative learning for reliable crowdsourcing systems. Proceedings of the 24th International Conference on Neural Information Processing Systems. 2011. pp. 1953–1961.

50. Zhang Y., Chen X., Zhou D., Jordan M.I. Spectral methods meet EM: A provably optimal algorithm for crowdsourcing. Proceedings of the 27th International Conference on Neural Information Processing Systems, NIPS'14, Advances in Neural Information Processing Systems. 2014. vol. 2. pp. 1260–1268.
51. Demartini G., Difallah D.E., Cudré-Mauroux P. ZenCrowd: leveraging probabilistic reasoning and crowdsourcing techniques for large-scale entity linking. Proceedings of the 21st international conference on World Wide Web (WWW '12). 2012. pp. 469–478.
52. Gupta M.R., Chen Y. Theory and Use of the EM Algorithm. Foundations and Trends in Signal Processing. 2010. vol. 4. no. 3. pp. 223–296.
53. McLachlan G.J., Krishnan T. The EM algorithm and extensions. 2nd ed. 2008. 359 p.
54. Sheshadri A., Lease M. Square: A benchmark for research on computing crowd consensus. Proceedings of the 1st AAAI Conference on Human Computations and Crowdsourcing. 2013. pp. 156–164.
55. Quoc Viet Hung N., Tam N.T., Tran L.N., Aberer K. An Evaluation of Aggregation Techniques in Crowdsourcing. Web Information Systems Engineering – WISE 2013, 2013. LNCS 8181. pp. 1–15.
56. Khosravifar B., Bentahar J., Gomrokchi M., Alam R. CRM: An efficient trust and reputation model for agent computing. *Knowledge-Based Systems*. 2012. vol. 30. pp. 1–16.
57. Fang H., Guo G., Zhang J. Multi-faceted trust and distrust prediction for recommender systems. *Decision Support Systems*. 2015. vol. 71. pp. 37–47.
58. Wahab O.A., Bentahar J., Otok H., Mourad A. A survey on trust and reputation models for Web services: Single, composite, and communities. *Decision Support Systems*. 2015. vol. 74. pp. 121–134.
59. Yu H. et al. Filtering trust opinions through reinforcement learning. *Decision Support Systems*. 2014. vol. 66. p. 102–113.
60. Liu S., Yu H., Miao C., Kot A.C. A fuzzy logic based reputation model against unfair ratings. Proceedings of the 2013 International Conference on Autonomous Agents and Multi-Agent Systems. 2013. pp. 821–828.
61. Auer P., Cesa-Bianchi N., Fischer P. Finite-time Analysis of the Multiarmed Bandit Problem. *Machine Learning*. 2002. vol. 47, no. 2/3. pp. 235–256.
62. Bubeck S., Cesa-Bianchi N. Regret analysis of stochastic and nonstochastic multi-armed bandit problems. 2012. 138 p.
63. Restuccia F., Das S.K., Payton J. Incentive Mechanisms for Participatory Sensing: Survey and Research Challenges. *ACM Transactions on Sensor Networks*. 2016. vol. 12. no. 2. Article no. 13.
64. Camerer C.F. Behavioral game theory: Predicting human behavior in strategic situations. *Advances in Behavioral Economics*. 2011. pp. 374–392.
65. Von Ahn L., Dabbish L. Labeling images with a computer game. Proceedings of Conference on Human Factors in Computing Systems. 2004. pp. 319–326.
66. Teng C.M. Correcting noisy data. Proceedings of the Sixteenth International Conference on Machine Learning. 1999. pp. 239–248.
67. Falikman M.V., Koifman A.Ya. [Types of priming in perception and perceptive attention research. Part 1]. *Vestnik Moskovskogo Universiteta. Seriya 14. Psihologiya – Herald of Moscow State University. Series 14. Psychology*. 2005. vol. 3. pp. 86–97. (In Russ.).
68. Frenay B., Verleysen M. Classification in the Presence of Label Noise: A Survey. *IEEE Transactions on Neural Networks and Learning Systems*. 2014. vol. 25. no. 5. pp. 845–869.
69. Eickhoff C., de Vries A.P. Increasing cheat robustness of crowdsourcing tasks. *Information Retrieval*. 2013. vol. 16. no. 2. pp. 121–137.