

Д. А. Усталов

ИСТОРИЯ, НАПРАВЛЕНИЯ И НЕКОТОРЫЕ ПРОБЛЕМЫ СОВРЕМЕННЫХ ИССЛЕДОВАНИЙ КРАУДСОРСИНГА КАК НАУЧНО-ПРАКТИЧЕСКОЙ ДИСЦИПЛИНЫ

Сегодня краудсорсинг является широко используемым способом решения многих задач сбора и агрегации данных. В данной работе проведен обзор исследований краудсорсинга как научно-практической дисциплины. Выделены направления исследований и сформулированы некоторые актуальные проблемы данной дисциплины.

Ключевые слова: *краудсорсинг, управление процессом, управление участниками, назначение заданий, оценка качества.*

Today, crowdsourcing became a popular approach for various data collecting and mining tasks. In this work, several modern crowdsourcing studies in different research trends have been discussed and some problems within these trends have been mentioned.

Keywords: *crowdsourcing, process control, worker management, task allocation, quality assessment.*

Введение

Сегодня *краудсорсинг* является широко используемым способом решения многих задач сбора и агрегации данных [1]. Примерами таких задач является обработка естественного языка, инженерия знаний, анализ данных, и др. Несмотря на это, до сих пор чрезвычайно актуальна проблема максимально эффективного использования труда заинтересованных участников. Их доступность и свободное время крайне ограничены, поэтому необходимо извлечь максимально возможную пользу от вклада каждого вовлеченного участника.

Исследования краудсорсинга

Целенаправленные исследования краудсорсинга начались в середине 2000-х гг., что совпало с появлением миллионной статьи в английской «Википедии» первого марта 2006 г. Это продемонстрировало возможность эффективного сбора и консолидации человеческого знания при помощи краудсорсинга в жанре «мудрость толпы». Исследователей краудсорсинга можно условно разделить на три временных периода, соответствующих теории диффузии инноваций Э. Роджерса [2]: «инноваторы» — конец 2000-х, «первопроходцы» — начало 2010-х, «раннее большинство» — середина 2010-х. На рис. 1 приведена классификация современных исследований краудсорсинга.

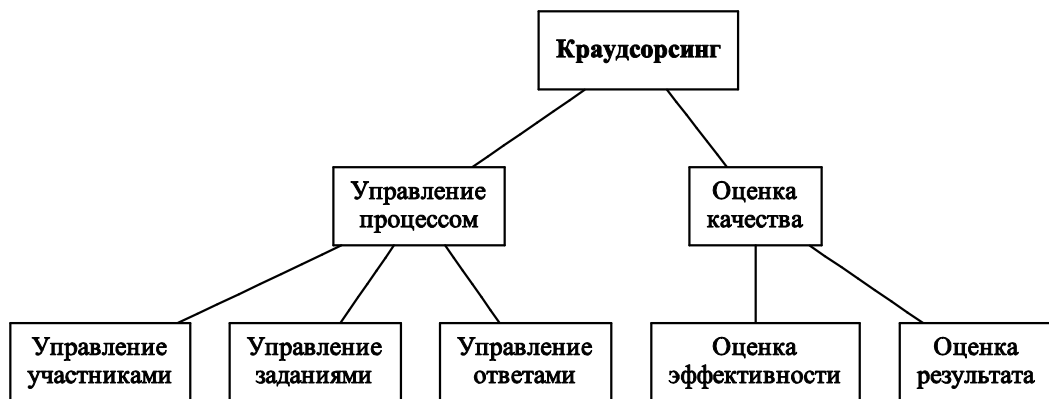


Рис. 1. Современные исследования краудсорсинга

Первоначальные исследования краудсорсинга в конце 2000-х годов были посвящены изучению его возможностей и ограничений с одной стороны, и попыток построить методы оценки качества созданных с его помощью ресурсов, с другой стороны. Исследователи пытались найти ответ на вопрос: «Работает ли краудсорсинг? Громко прозвучавшие истории успеха таких проектов, как «Википедия», Amazon Mechanical Turk (MTurk) и iStockphoto, дали положительный ответ на этот вопрос [1].

Ближе к концу 2000-х гг. на коммерческих биржах краудсорсинга возникла острая проблема спама — случайных или бессмысленных ответов, отправленных участниками в мошеннических целях. Участники таких бирж получают небольшие деньги за выполнение достаточно простых заданий, решение которых требует несколько минут. Количество полученных участниками денег напрямую зависит от количества выполненных заданий — это привело к тому, что в 2011 г. спамеры составляли 39 % пользователей платформы Amazon Mechanical Turk [3]. Исследования вопроса планирования процесса, проблем спама в краудсорсинге и доверия к неконтролируемой толпе участников были важной темой работ начала 2010-х годов с доминирующим вопросом: «Как сделать краудсорсинг надёжнее?»

С появлением различных способов оценки репутации участников и выявления спамеров, в 2010-х гг. возник закономерный интерес к повышению эффективности процесса краудсорсинга в самых разных смыслах этого слова. Большое внимание начало уделяться задачам генерации заданий по участникам, проблемам агрегации надежных и непротиворечивых ответов, вопросам кооперации, геймификации и общей доступности процесса [4]. Сегодня доминирующим является следующий вопрос: как сделать краудсорсинг дешевле?

В настоящее время краудсорсинг является важной научно-практической дисциплиной, которой посвящены отдельные конференции и семинары, например CSCW при поддержке Ассоциации вычислительной техники, HCOMP при поддержке Ассоциации развития искусственного интеллекта, GamifIR в рамках Европейской конференции по информационному поиску, и др.

Управление краудсорсингом

Исследования вопросов управления процессом краудсорсинга посвящены построению типовых схем краудсорсинга для решения распространенных практических задач.

Ранние работы в этой области относились преимущественно к задачам машинного зрения и разметки изображений при помощи тегов. Одной из первых и наиболее цитируемых работ является статья Л. фон Ана и Л. Дэббис из Университета Карнеги-Меллон (США) об игрофицированном способе разметки изображений путем согласования одновременных ответов игроков [5].

В 2009 году, Дж. Уайтхилл и группа исследователей из Калифорнийского университета в Сан-Диего (США) представили модель *GLAD*, позволяющую объединить в единую систему такие параметры, как сложность назначенных заданий, квалификацию участников, и получаемые ответы [6].

В 2010 г., М. Бернштейн и группа исследователей из нескольких университетов США представили текстовый процессор *Soylent*, являющийся тесно интегрированной с платформой Amazon Mechanical Turk надстройкой над редактором Microsoft Word. *Soylent* позволяет прозрачным образом привлекать участников платформы с целью форматирования, перефразирования и доработки текста [7]. Важным результатом данной работы является подход *Find-Fix-Verify*, разделяющий задания с открытым вопросом на независимые части.

В 2012 г. Г. Демартини, Дж. Дифаллах и Ф. Кудре-Мару из Университета Фрибура (Швейцария) разработали модель *ZenCrowd* [8], изначально предназначенную для крупномасштабной разметки ссылок новостей при помощи краудсорсинга на MTurk. Модель предполагает автоматическую предварительную обработку данных и применяет вероятностную сеть для агрегации ответов участников.

Управление участниками

В начале 2010-х г. стало ясно, что далеко не все участники процесса краудсорсинга настроены на добросовестное выполнение работы. Возникли серьезные проблемы со спамом и просто неопытными участниками, вносящими ошибки в результаты. Основные направления исследований в области управления участниками в краудсорсинге — борьба с мошенничеством, обучение пользователей и оценка их компетентности.

П. Велиндер и П. Перона из Калифорнийского технологического института (США) предложили интерактивный метод оценки характеристик участников разметки на основе вероятностной модели [9]. Применение этого метода в задаче обнаружения границ изображений позволило снизить количество необходимых оценок для разметки и уменьшить долю допускаемых участниками ошибок.

Работа Х. Йу и соавторов из Наньянского технологического университета (Сингапур) и Китайской академии наук посвящена успешному рисованию изображения сто долларововой купюры при помощи краудсорсинга путём сбора множества

нарисованных фрагментов этой купюры [10]. В работе представлен метод ранжирования участников по уровню доверия.

Алгоритм *WorkerRank*, созданный М. Дальтаяни, Л. де Альфаро и П. Пападимитроу из Калифорнийского университета в Санта-Крузе и компании Elance-oDesk (США), позволяет одновременно оценить репутацию участников и относительную важность заданий с учётом заранее заявленных компетенций [11]. Представленный алгоритм предназначен для рекомендации исполнителей заданий в области информационных технологий и показывает лучший результат, чем методы коллаборативной фильтрации.

Управление заданиями

Исследователи-первопроходцы в области краудсорсинга уделяют большое внимание механизмам управления заданиями, связанными с их распределением и оценкой сложности выполнения. Несмотря на то, что в общем виде задача по перераспределению заданий по участникам является трудноразрешимой [12], с 2011 г. В этой области возник заметный и закономерный интерес.

С. Ли, С. Парк и С. Парк из Университета Соган (Южная Корея) создали метод динамического программирования, позволяющий выполнить распределение заданий по пользователям [13]. Эксперименты показывают, что учёт уровня квалификации участника и сложности задания позволяет повысить качество результата.

В том же году Д. Каргер, С. Ох и Д. Шах из Массачусетского технического института и Иллинойского университета в Урбане-Шампейне (США) предложили алгоритм вывода назначения заданий, оптимальный по порядку [14]. Важным выводом из этой работы является гипотеза о *фазовом переходе*, показывающая невозможность улучшить качество метода голосования большинства при низком количестве заявок на выполнение задания.

Алгоритмы *BudgetFix* и *Budgeteer*, разработанные Л. Тран-Тханом и соавторами из Саутгемптонского университета, позволяют выполнить назначение заданий при заданных ограничениях бюджета с гарантией выбранной аккуратности [15]. Данные методы работают эффективнее, чем популярный метод *Soylent* (см. [7]).

Управление ответами

Р. Сноу и ряд других известных исследователей из Стэнфордского университета в 2008 г. показали, что при построении лингвистических ресурсов для достижения качества краудсорсинга на уровне профессиональных лингвистов достаточно *четырёх* ответов на каждый вопрос [16]. При этом в работе предложена вероятностная модель, позволяющая снизить смещенность ответов, получаемых от участников.

Б. Ханрахан, Г. Конвертино и Л. Нельсон из Xerox PARC (США) провели исследование популярного вопрос-ответного сервиса по информационным технологиям *Stack Overflow* и предложили метод оценки сложности вопросов на основе

коэффициента корреляции Пирсона и метод оценки уровня квалификации участников на основе меры Z -score [17].

М. Йоглекар, Г. Гарсия-Молина и А. Парамесваран из Стэнфордского университета в том же году представили метод построения доверительного интервала для оценки доли ошибок участников процесса краудсорсинга [18]. Данный метод позволяет также построить доверительный интервал для аккуратности участников.

Оценка качества

Краудсорсинг позволяет выполнять относительно недорогую разметку различных данных, сравнимую по качеству с работой приглашенных экспертов. Это подтверждается исследованием Дж. Хира и М. Бостока из Стэнфордского университета, оценившим шестикратную разницу в стоимости в пользу краудсорсинга для задачи оценки изображений [19].

В 2011 г. Д. Олесон и коллеги из компании CrowdFlower представили метод для масштабируемого контроля качества краудсорсинга путём направленного предоставления обратной связи об ответах участников, что привело к сокращению ручной работы при управлении качеством результата [20].

Г. Зоу, А. Гил и М. Тхарайил из Даляньского технологического университета (Китай) и Исследовательского центра Xerox в Пало-Альто (США) в 2014 г. выполнили агентное моделирование поведения участников процесса краудсорсинга [21]. В работе предложена обученная модель на основе линейной регрессии и логнормального распределения, позволяющая оценить аккуратность выполнения задания участником.

Результаты и выводы

Табл. 1, составленная на основе проведенного обзора, иллюстрирует направления для дальнейших исследований. В целях экономии места, используются следующие обозначения областей исследований: УП — управление процессом, УУ — участниками, УЗ — заданиями, УО — ответами, ОК — оценка качества.

Таблица 1

Некоторые проблемы краудсорсинга

№	Область	Проблема	Ссылки
1	УП	Отсутствие открытого программного обеспечения для управления процессом.	[7, 15]
2	УУ	Оценка квалификации участника по классу заданий.	[11, 21]
3	УЗ	Интерактивная генерация графа назначения заданий.	[14, 15]
4	УО	Оценка сложности заданий на основе ответов участников.	[13, 17]
5	УП+ОК	Назначение заданий участникам с учетом сложности, квалификации, бюджета.	[11, 14, 15, 17]

Заключение

В данной работе проведен обзор исследований краудсорсинга как научно-практической дисциплины. Выделены направления исследований и сформулированы некоторые актуальные проблемы в рамках этих направлений.

Благодарность

Работа поддержана грантом РГНФ № 13-04-12020 «Новый открытый электронный тезаурус русского языка».

Литература

1. *Estellés-Arolas E., González-Ladrón-de Guevara F.* Towards an integrated crowdsourcing definition // *Journal of Information Science*. 2012. Vol. 38. No. 2. P. 189–200.
2. *Rogers E.* Diffusion of innovations. 4th edition. Simon and Schuster, 2010. 518 pp.
3. *Vuurens J. B. P., de Vries A. P., Eickhoff C.* How Much Spam Can You Take? An Analysis of Crowdsourcing Results to Increase Accuracy // *Proc. ACM SIGIR Workshop on Crowdsourcing for Information Retrieval (CIR'11)*. 2011. Pp. 21–26.
4. *Ustalov D.* Towards Crowdsourcing and Cooperation in Linguistic Resources // *Proceedings of RuSSIR 2014 and the associated Young Scientist Conference* / Ed. by P. Braslavskiy et al. 2014. *Communications in Computer and Information Science*.
5. *Von Ahn L., Dabbish L.* Labeling Images with a Computer Game // *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. CHI '04*. New York, NY, USA: ACM, 2004. Pp. 319–326.
6. *Whitehill J., Ruvolo P., Wu T. et al.* Whose Vote Should Count More: Optimal Integration of Labels from Labelers of Unknown Expertise // *Advances in Neural Information Processing Systems 22*. Curran Associates, Inc., 2009. Pp. 2035–2043.
7. *Bernstein M. S., Little G., Miller R. C. et al.* Soylent: A Word Processor with a Crowd Inside // *Proceedings of the 23rd Annual ACM Symposium on User Interface Software and Technology. UIST '10*. New York, NY, USA: ACM, 2010. Pp. 313–322.
8. *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*. New York, NY, USA: ACM, 2012. Pp. 469–478.
9. *Welinder P., Perona P.* Online Crowdsourcing: Rating Annotators and Obtaining Cost-Effective Labels // *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. 2010. Pp. 25–32.
10. *Yu H., Shen Z., Miao C., An B.* Challenges and Opportunities for Trust Management in Crowdsourcing // *Proceedings of the The 2012 IEEE/WIC/ACM International Joint Conferences on Web Intelligence and Intelligent Agent Technology — Volume 02. WI-IAT '12*. Washington, DC, USA: IEEE Computer Society, 2012. Pp. 486–493.
11. *Daltayanni M., de Alfaro L., Papadimitriou P.* WorkerRank: Using Employer Implicit Judgments to Infer Worker Reputation // *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining. WSDM '15*. New York, NY, USA: ACM, 2015. Pp. 263–272.
12. *Gujar S., Narahari Y.* Redistribution Mechanisms for Assignment of Heterogeneous Objects // *Journal of Artificial Intelligence Research*. 2011. Vol. 41. Pp. 131–154.

13. Lee S., Park S., Park S. A Quality Enhancement of Crowdsourcing based on Quality Evaluation and User-Level Task Assignment Framework // 2014 International Conference on Big Data and Smart Computing (BIGCOMP). IEEE, 2014. Pp. 60–65.
14. Karger D. R., Oh S., Shah D. Budget-Optimal Task Allocation for Reliable Crowdsourcing Systems // Operations Research. 2014. Vol. 62, no. 1. Pp. 1–24.
15. Tran-Thanh L., Huynh T. D., Rosenfeld A. et al. Crowdsourcing Complex Workflows under Budget Constraints // Twenty-Ninth AAAI Conference on Artificial Intelligence (AAAI-15). AAAI Press, 2015.
16. Snow R., O'Connor B., Jurafsky D., Ng A. Y. Cheap and Fast—but is It Good?: Evaluating Non-expert Annotations for Natural Language Tasks // Proceedings of the Conference on Empirical Methods in Natural Language Processing. EMNLP '08. Stroudsburg, PA, USA: Association for Computational Linguistics, 2008. Pp. 254–263.
17. Hanrahan B. V., Convertino G., Nelson L. Modeling Problem Difficulty and Expertise in StackOverflow // Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work Companion. CSCW '12. New York, NY, USA: ACM, 2012. Pp. 91–94.
18. Joglekar M., Garcia-Molina H., Parameswaran A. Evaluating the Crowd with Confidence // Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. KDD '13. New York, NY, USA: ACM, 2013. Pp. 686–694.
19. Heer J., Bostock M. Crowdsourcing Graphical Perception: Using Mechanical Turk to Assess Visualization Design // Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. CHI '10. New York, NY, USA: ACM, 2010. Pp. 203–212.
20. Oleson D., Sorokin A., Laughlin G. P. et al. Programmatic Gold: Targeted and Scalable Quality Assurance in Crowdsourcing // Human Computation: Papers from the 2011 AAAI Workshop (WS-11-11). 2011. Pp. 43–48.
21. Zou G., Gil A., Tharayil M. An Agent-based Model for Crowdsourcing Systems // Proceedings of the 2014 Winter Simulation Conference. WSC '14. Piscataway, NJ, USA: IEEE Press, 2014. Pp. 407–418.