The Ever Evolving Online Labor Market: Overview, Challenges and Opportunities

Sihem Amer-Yahia[◇],Senjuti Basu Roy[†].
[◇] CNRS, Univ. Grenoble Alpes, [†]NJIT
sihem.amer-yahia@cnrs.fr,senjutib@njit.edu

ABSTRACT

The goal of this tutorial is to make the audience aware of various discipline-specific research activities that could be characterized to be part of online labor markets and advocate for a unified framework that is interdisciplinary in nature and requires convergence of different research disciplines. We will discuss how such a framework could bring transformative effect on the nexus of humans, technology, and the future of work.

PVLDB Reference Format:
Sihem Amer-Yahia, Senjuti Basu Roy. The Ever Evolving Online Labor Market: Overview, Challenges and Opportunities. *PVLDB*, 12(12): 1978-1981, 2019.
DOI: https://doi.org/10.14778/3352063.3352114

1. OVERVIEW AND RELEVANCE

The rapid development of professional social networks and online labor markets, is affecting the future of jobs and workers. Professional social networks such as LinkedIn, are revolutionizing hiring practices. An increasing number of individuals rely on such networks to find jobs, and it is becoming common practice for head hunters and companies to examine one's profile on LinkedIn before contacting or hiring someone. Online job marketplaces are gaining popularity as mediums to hire people to perform certain tasks. These marketplaces include freelancing platforms such as Qapa and MisterTemp' in France, and TaskRabbit and Fiverr in the USA. On those platforms, workers can find temporary jobs in the physical world (e.g., looking for a plumber), or in the form of virtual "micro-gigs" such as "help with HTML, JavaScript, CSS, and JQuery". Crowdsourcing platforms are a very popular type of online job marketplaces nowadays. These platforms are fully virtual: workers are hired online and tasks are also completed online. Examples of crowdsourcing platforms are FouleFactory, and Prolific Academic in Europe, and Amazon Mechanical Turk and Figure Eight in the USA. As the gig economy grows, an important offshoot of this movement is "flash organizations", the pop-up shops that come together for a purpose, such as developing and marketing an app. They include entire teams that work closely together, and when the project is complete, the teams disband. For example, Gigster.com is a smart development service that brings freelancers together into software-building teams on demand at any time, or Artella.com is designed for creating complex animated features, offering teams of freelance animators, sound designers, and other talent at your beck and call.

From a conceptual standpoint, online labor markets can be characterized as follows: workers are the producers of work that are to be consumed by systems, platforms, or organizations. Workers need to be hired, treated, and compensated fairly, they have to be assisted to accomplish their work or further their career. On the other hand, consumers, such as, business, organizations, or platforms need to hire an appropriate workforce to be able to accomplish their business goal in time, with certain accuracy, and within budget.

Social science as well as data-centric research [5, 7, 8, 10, 13, 14, 18, 23, 26, 27] have developed disconnected and discipline-specific approaches to solve different problems in online labor markets. For example, social science researchers have proposed conceptual frameworks, visualization, and software prototypes to recruit workers and form appropriate teams, decompose complex tasks to assist workers, or propose tools to enable interactivity and collaboration among workers. Database research, on the other hand, has addressed scalability and data management challenges of how to curate and clean data produced by online labor markets, how to model the data and store it effectively, or how to form appropriate teams [3, 21, 22]. Machine learning research [7, 14, 27] proposed models that are capable of aggregating workers' contributions or infer/estimate groundtruths when that is unknown. Finally, psychology and organizational research [17, 25] proposed conceptual models that are required to understand and analyze human behavior in online labor market.

We will first describe different applications that rely on free-lancing and online labor markets. Then, we will re-visit seminal and prominent works that came out from different research communities (social science, machine learning, data management, psychology and organizational research, and economics) on the topic of online labor markets, describe their approaches and summarize their impact on science, society and industry. Finally, we will outline the requirements of a unified framework that has the potential to combine the best of all these worlds and present modeling, data management, and algorithmic challenges to conceptualize it.

Proceedings of the VLDB Endowment, Vol. 12, No. 12
ISSN 2150-8097.
DOI: https://doi.org/10.14778/3352063.3352114

2. SCOPE, DURATION AND STRUCTURE

The tutorial aims to gather existing work in several research areas and understand the new challenges and opportunities in the future of work, that are of interest to the data management community. It is intended for **1.5 hours** and is organized into three parts:

PART I: Applications (15 minutes) In this part, we will describe different applications that tap into on-line labor markets. The applications range from free-lancing, crowdsourcing, citizen science, as well as flash organizations. We will characterize these applications by describing the nature of the type of work/business and workers, thereby highlighting the desirable properties that each must meet. In particular, we will characterize applications that are micro-gig/macro-gig, ones that require collaboration/could be completed independently, and ones that require machine participation or human-only.

PART II: Existing Approaches (50 minutes) This piece of the tutorial will revisit existing efforts and summarize them along three dimensions.

- *Data and Problem Modeling (20mn):* Online platforms are bringing transformational changes in the process of recruiting and retaining the workforce and the nature of work. Either the work is fully online (e.g., hiring someone online to design a website), fully offline (e.g., relying on reference letters on LinkedIn to hire a person), or hybrid (e.g., using TaskRabbit to hire help to move furniture). Jobs or work, on the other hand, are becoming more "fluid" and [1] on-demand, and workers need to be dynamically assigned to such jobs. In that context, the workforce is becoming volatile and with varying expertise levels. The reliance on algorithms to match workers and jobs is today a reality. Such algorithms rely on a data model and are designed to solve specific problems. The first part of the tutorial will review existing platforms and characterize them in terms of their data and problem modeling. We will discuss how each of the aforementioned research communities, i.e., social science, data management, machine learning, psychology and organizational studies have modeled online labor market problems.

  For example, the traditional psychology or organizational management research have proposed conceptual models grounded on human psychology or socio-cognitive behavior that empirically studies human performance based on motivation, fun, monotony, etc. We will study how social science [12, 23, 16, 18] and psychology research have addressed human factors [2] modeling, for the problems of incentive design, task decomposition, task assignment, worker engagement, and retention. On the contrary, even though these factors are recognized to be important in machine learning and data management research [24], to the best of our knowledge, only a handful of existing works[20], consider these factors in problem model or algorithm design. In fact, these data centric communities put more focus in proposing models that are more platform specific (such as quality, cost, latency) leading to computationally rigorous models. Unfortunately, humans or producers of works do not receive that much of attention, primarily because human behavior modeling requires expertise in psychology and social science. Our objective will be to discuss these issues and do a deep dive on their proposed modeling techniques.

- *Solution Techniques (20mn):* Computational social scientists, theoreticians, data management and data mining researchers, have proposed different approaches that solve the worker recruitment, job selection, worker engagement, team formation and learning [1, 15], as well as skill estimation problems [11]. Additionally, the design of platforms raises a number of challenges that have been addressed in different research communities.

  For example, Social Science research has taken a semi-computational approach that business and organizations could make use of. Recent works have also studied automated approaches (algorithms) for workers recruitment, team formation,and task decomposition. Nevertheless Social Science researchers have spent significant effort in designing empirical and field studies that validate their hypotheses. They have also developed visualizations and software to hire, retain, and augment the online workforce. The database and machine learning communities, on the other hand, have taken fully computational approaches. In fact, the common practice in machine learning is to propose some models that are suitable for the underlying research problem. For example, task recommendation has been modeled as multi-arm bandit and as matrix factorization problems, and truth discovery, result aggregation, skill learning were studied as supervised learning problems where EM (Expectation Maximization) types of solutions were designed. Such solutions aim to produce maximum-likelihood estimates of parameters when there is a many-to-one mapping from an underlying distribution to the distribution governing the observation. The data management community have studied most of the online labor market problems as discrete optimization problems and emphasized scalability and algorithmic challenges. We intend to review and summarize these different solution techniques.

- *Impact (10mn):* We end this part with a review of the impact of existing works. Impact will be characterized from both scientific, societal and industry viewpoints.

  - Scientific impact: computational solutions have been designed to address learning and augmentation, feature engineering, inferring skills [22], and learning motivation functions [19].
  - Societal impact: empirical and computational approaches have been designed to address human capital advancement, peer learning and distributed mentoring, as well as adaptive worker-centric solutions to account for evolving worker motivation. Additional approaches have been proposed to study transparency, fairness and privacy in online job markets [6, 9].
  - Industrial impact: platforms and software that came out recently, such as platform UpWork, or

[1] https://www.nytimes.com/2017/07/12/business/economy/flash-organizations-labor.html

the True Story software[2], that hired freelance designers and writers to develop games.

PART III: Toward a Unified Framework (25 minutes) This part is primarily forward looking and aims to discuss the challenges and opportunities that arise from bringing together empirical and computational approaches to unify the design of online labor markets. We will look beyond and describe our take on how online labor markets can contribute to the growth of landscape of work in future.

The two big challenges we envision are: capturing humans and their changing and unpredictable nature, and designing approaches that address the variety of goals underlying job marketplaces. This will require us to have the ability to process large amounts of data on jobs and workers in a scalable and fair way. These two challenges raise new opportunities that will be discussed as open research questions in this part of the tutorial.

- *Modeling Opportunities:* Open questions include how to model humans with different roles and intents (perfecting skills, earning money, communicating with others), how to model teams, and how to make the model evolve over time). We will discuss the use of clustering and active learning approaches to capture context and different roles. We will also discuss the role of indexing in managing different human models and their evolution.

- *Optimization Opportunities:* Open questions include how to best express and leverage optimization-based approaches that blend human-centric and platform-centric goals. We will discuss the need for multi-objective optimization approaches, inspired from multi-objective query optimization and skyline queries, to navigate in the space of solutions in a scalable fashion. We will discuss the relevance of this proposal to optimize task matching, multi-stakeholders fairness, and peer learning in the presence of different affinities between workers.

3. TARGET AUDIENCE, PREREQUISITES

The tutorial will be of interest to both theoreticians and practitioners who are interested in the development of novel data-centric applications in the areas of databases, data mining, machine learning, social science, and algorithms, ranging from large-scale analytics to emerging online applications. Tutorial attendees are expected to have basic knowledge in machine learning, algorithms, and data management. Knowledge in constrained optimization is not necessary.

4. RELEVANCE

The proposed tutorial is timely as it addresses unsolved questions in the emerging area of the future of work. The tutorial is relevant to the general area of data management and the web and more specifically, to Big Data Processing and Transformation, Data Mining, Clustering and Knowledge Discovery, Large-Scale Analytics, Indexing, Query Processing and Optimization, Social Networks Analysis, Graph Databases, Information Retrieval. The technical topics covered are constrained optimization, hardness results, ranking semantics and fairness, algorithms, and empirical evaluations.

The authors have published seminal papers on human modeling and scalable task assignment in crowdsourcing [3, 19], on team formation for and human factor estimation [21, 22], and on fairness in virtual marketplaces [4].

5. RELATED TUTORIALS

While there have been a few tutorials on crowdsourcing, and most notably the one mentioned below, those tutorials do not address the convergence of multiple disciplines for the future of work or data management and scalability questions in relation to modeling humans.

- Yongxin Tong, Lei Chen, Cyrus Shahabi: Spatial Crowdsourcing: Challenges, Techniques, and Applications. PVLDB 10(12): 1988-1991 (2017)

- Sihem Amer-Yahia, Senjuti Basu Roy: Human Factors in Crowdsourcing. PVLDB 9(13): 1615-1618 (2016)

6. BIOGRAPHY

Sihem Amer-Yahia is a CNRS Research Director at the University of Grenoble Alpes where she leads the SLIDE team. Her interests are at in large-scale data management. Before joining CNRS, she was Principal Scientist at the Qatar Computing Research Institute, Senior Scientist at Yahoo! Research and at&t Labs. Sihem has served on the SIGMOD Executive Board, is a member of the VLDB and the EDBT Endowments. She is the Editor-in-Chief of the VLDB Journal and an Associate Editor for Transctions in Data Science. She was PC chair of VLDB 2018. Sihem received her Ph.D. in Computer Science from Paris-Orsay and INRIA in 1999, and her Diplôme d'Ingénieur from INI, Algeria. Sihem is co-organizing a Shonan Meeting on the topic of Human-in-the-loop Big Data and AI: Connecting Theories and Practices for a Better Future of Work.

Senjuti Basu Roy is an Assistant Professor at the New Jersey Institute of Technology. Senjutis broader research interests lie in the area of data andcontent management of web and structured data with a focus on ex-ploration, analytics, and algorithms. In recent years, her research hasfocused on designing principled algorithms and systems that requireman-machine collaboration. Senjuti has presented three tutorials onthe computational challenges related to man-machine systems inVery Large Database Conference, International World Wide WebConference, and International Conference on Extending DatabaseTechnology. She was the PC Co-chair of SIGMOD 2018 mentorshiptrack, and the PC co-chair of VLDB 2018 PhD Workshop program. Senjuti was a co-organizer of ExploreDB 2016 (co-located with SIGMOD 2016) and the IEEE Workshop on Human-in-the-loop Methods and Human Machine Collaboration in BigData (IEEE HMData2017, 2018, 2019) (co-located with IEEE Big data). She has organized an NSF workshop on converging human and technological perspectives in crowdsourcing research and will be co-organizing a Shonan Meeting on the topic of Human-in-the-loop Big Data and AI: Connecting Theories and Practices for a Better Future of Work.

[2]https://qz.com/1027606/forget-the-on-demand-worker-stanford-researchers-built-an-entire-on-demand-organization/

Acknowledgement

The work of Senjuti Basu Roy is supported by National Science Foundation, Grant No. 1814595 and Office of Naval Research, grant no: N000141812838.

7. REFERENCES

[1] R. Agrawal, B. Golshan, and E. Terzi. Grouping students in educational settings. In *SIGKDD*, pages 1017–1026. ACM, 2014.

[2] S. Amer-Yahia and S. B. Roy. Human factors in crowdsourcing. *PVLDB*, 9(13):1615–1618, 2016.

[3] S. Basu Roy, I. Lykourentzou, S. Thirumuruganathan, S. Amer-Yahia, and G. Das. Task assignment optimization in knowledge-intensive crowdsourcing. *The VLDB JournalThe International Journal on Very Large Data Bases*, 24(4):467–491, 2015.

[4] R. M. Borromeo, T. Laurent, M. Toyama, and S. Amer-Yahia. Fairness and transparency in crowdsourcing. In *EDBT*, pages 466–469, 2017.

[5] F. Daniel, P. Kucherbaev, C. Cappiello, B. Benatallah, and M. Allahbakhsh. Quality control in crowdsourcing: A survey of quality attributes, assessment techniques, and assurance actions. *ACM Computing Surveys (CSUR)*, 51(1):7, 2018.

[6] S. Elbassuoni, S. Amer-Yahia, C. E. Atie, A. Ghizzawi, and B. Oualha. Exploring fairness of ranking in online job marketplaces. In *EDBT (short paper)*, 2019.

[7] J. Gao, Q. Li, B. Zhao, W. Fan, and J. Han. Truth discovery and crowdsourcing aggregation: A unified perspective. *PVLDB*, 8(12):2048–2049, 2015.

[8] H. Garcia-Molina, M. Joglekar, A. Marcus, A. Parameswaran, and V. Verroios. Challenges in data crowdsourcing. *IEEE TKDE*, 28(4):901–911, 2016.

[9] A. Ghizzawi, J. Marinescu, S. Elbassuoni, S. Amer-Yahia, and G. Bisson. Fairank: An interactive system to explore fairness of ranking in online job marketplaces. In *EDBT (demonstration)*, 2019.

[10] D. Haas, J. Ansel, L. Gu, and A. Marcus. Argonaut: macrotask crowdsourcing for complex data processing. *PVLDB*, 8(12):1642–1653, 2015.

[11] M. Joglekar, H. Garcia-Molina, and A. Parameswaran. Evaluating the crowd with confidence. In *SIGKDD*, pages 686–694. ACM, 2013.

[12] N. Kaufmann, T. Schulze, and D. Veit. More than fun and money. worker motivation in crowdsourcing-a study on mechanical turk. In *AMCIS*, volume 11, pages 1–11. Detroit, Michigan, USA, 2011.

[13] W. S. Lasecki, J. Kim, N. Rafter, O. Sen, J. P. Bigham, and M. S. Bernstein. Apparition: Crowdsourced user interfaces that come to life as you sketch them. In *CHI*, pages 1925–1934. ACM, 2015.

[14] C. H. Lin, E. Kamar, and E. Horvitz. Signals in the silence: Models of implicit feedback in a recommendation system for crowdsourcing. In *AAAI*, 2014.

[15] Y. Liu, Q. Liu, R. Wu, E. Chen, Y. Su, Z. Chen, and G. Hu. Collaborative learning team formation: a cognitive modeling perspective. In *DSAA*, pages 383–400. Springer, 2016.

[16] A. Mao, E. Kamar, Y. Chen, E. Horvitz, M. E. Schwamb, C. J. Lintott, and A. M. Smith. Volunteering versus work for pay: Incentives and tradeoffs in crowdsourcing. In *First AAAI conference on human computation and crowdsourcing*, 2013.

[17] S. Mohammed, R. Klimoski, and J. R. Rentsch. The measurement of team mental models: We have no shared schema. *Organizational Research Methods*, 3(2):123–165, 2000.

[18] M. Nebeling, A. To, A. Guo, A. A. de Freitas, J. Teevan, S. P. Dow, and J. P. Bigham. Wearwrite: Crowd-assisted writing from smartwatches. In *CHI*, pages 3834–3846. ACM, 2016.

[19] J. Pilourdault, S. Amer-Yahia, D. Lee, and S. B. Roy. Motivation-aware task assignment in crowdsourcing. In *EDBT*, pages 246–257, 2017.

[20] J. Pilourdault, S. Amer-Yahia, S. B. Roy, and D. Lee. Task relevance and diversity as worker motivation in crowdsourcing. In *2018 IEEE 34th International Conference on Data Engineering (ICDE)*, pages 365–376. IEEE, 2018.

[21] H. Rahman, S. B. Roy, S. Thirumuruganathan, S. Amer-Yahia, and G. Das. Optimized group formation for solving collaborative tasks. *VLDB Journal*, 28(1):1–23, 2019.

[22] H. Rahman, S. Thirumuruganathan, S. B. Roy, S. Amer-Yahia, and G. Das. Worker skill estimation in team-based tasks. *PVLDB*, 8(11):1142–1153, 2015.

[23] D. Retelny, S. Robaszkiewicz, A. To, W. S. Lasecki, J. Patel, N. Rahmati, T. Doshi, M. Valentine, and M. S. Bernstein. Expert crowdsourcing with flash teams. In *UIST*, pages 75–85. ACM, 2014.

[24] S. B. Roy, I. Lykourentzou, S. Thirumuruganathan, S. Amer-Yahia, and G. Das. Crowds, not drones: modeling human factors in interactive crowdsourcing. In *DBCrowd 2013-VLDB Workshop on Databases and Crowdsourcing*, pages 39–42. CEUR-WS, 2013.

[25] R. Silberzahn and E. L. Uhlmann. Crowdsourced research: Many hands make tight work. *Nature News*, 526(7572):189, 2015.

[26] M. A. Valentine, D. Retelny, A. To, N. Rahmati, T. Doshi, and M. S. Bernstein. Flash organizations: Crowdsourcing complex work by structuring crowds as organizations. In *CHI*, pages 3523–3537. ACM, 2017.

[27] N. Wang and D.-Y. Yeung. Ensemble-based tracking: Aggregating crowdsourced structured time series data. In *ICML*, pages 1107–1115, 2014.