From Complex Object Exploration to Complex Crowdsourcing

Sihem Amer-Yahia
CNRS/LIG, France
Sihem.Amer-Yahia@imag.fr

Senjuti Basu Roy
UW Tacoma, USA
senjutib@uw.edu

1. ABSTRACT

Forming and exploring complex objects is at the heart of a variety of emerging web applications. Historically, existing work on complex objects has been developed in two separate areas: composite item retrieval and team formation. At the same time, emerging applications that harness the wisdom of crowd workers, such as, document editing by workers, sentence translation by fans (or fan-subbing), innovative design, citizen science or journalism, represent complex crowdsourcing, in which an object may represent a complex task formed by a set of sub-tasks or a team of workers who work together to solve the task. The goal of this tutorial is to bridge the gap between composite item retrieval and team formation and define new research directions for complex crowdsourcing applications.

2. GOAL

Composite item retrieval is prevalent in several web applications, such as online shopping, where products are bundled together to provide discounts, or travel itinerary recommendation, where points of interest in a city are combined into a single trip offer. Team formation problems are prevalent in the area of social networks analysis and group recommendations, where the objective is to form a team to solve a problem, or consume some items together. Both composite item retrieval and team formation problems are expressed as constrained optimization problem in the literature. Indeed, package retrieval must satisfy constraints (e.g., an iPhone with compatible accessories) and optimize an objective (e.g., have a low total price, or maximize the relevance, or both). Similarly, in group recommendation, constraints may represent the type of items to recommend and the objective could be to minimize disagreement between group members, and or maximize relevance.

While most crowdsourcing platforms are based on simple micro-tasks, such as, rating a movie or recognizing elements in a picture, complex crowdsourcing (CC) on the other hand is acknowledged as one of the most promising areas of next-generation crowdsourcing. CC are collaborative many times, and require workers to create knowledge content (for example, Wikipedia articles, or news articles) together through crowdsourcing. Crowd workers, each having a certain degree of expertise, collaborate and build on each other's contributions to gradually increase the quality of each knowledge piece (hereby referred to as task). For the example of fan-subbing, an objective may be to maximize the quality of the translated movie, while keeping the cost under the budget, or balancing the workload of the workers (constraints). Another rising application for complex tasks is reporting a crisis. Systems, such as CrowdMap, allow geographically closed people (optimization objective) with complementary skills (constraint), to work together to report details about the course of a typhoon or the aftermath of an earthquake. For all the scenarios described above, in order for CC applications to work effectively, worker teams have to be formed to complete tasks. Such a requirement would greatly benefit from formalisms and algorithms from both composite item retrieval, to specify task composition, and team formation, to identify sets of workers who can collaborate to complete a complex task. The formalisms and algorithms of composite item retrieval will guide *passive crowdsourcing* scenarios, such as citizen science, where citizen science volunteers are present and available as inputs (and can not be changed), but the objective is to form the complex tasks to exploit the expertise of the available workers as much as possible. On the other hand, *active crowdsourcing scenarios*, such as fan subbing or document editing will benefit from team formation work, where, given the set of tasks and available worker pool, the best set of workers is to be selected to assign to the tasks to ensure best outcome.

This tutorial will review the literature and research advances in the areas of composite item retrieval and of team formation. The main objective is to draw connections between two research areas that evolved separately and understand how the applications, formalizations, algorithms and empirical results obtained in each area can be used to solve emerging crowdsourcing applications.

3. TOPIC AND DESCRIPTION

Today's web applications manipulate complex objects ranging from package retrieval to ad-hoc team formation. The proposed tutorial aims to gather existing work in these areas and understand the new challenges and opportunities in emerging web applications. We will start with a review of a number of web applications that have been developed in practice in the last few years and evolve into a summary

WWW 2015 Companion, May 18–22, 2015, Florence, Italy.
ACM 978-1-4503-3473-0/15/05.
http://dx.doi.org/10.1145/2740908.2741994

of the various formalizations proposed to solve complex object formation in those applications. It will then focus on the algorithmic challenges and solutions as well as a summary of the empirical findings. The last part of the tutorial will describe the future opportunities and directions.

3.1 Overview of Existing Applications

The first part of this tutorial (30mn) will describe a variety of applications that share the principle of optimization-guided complex object formation, either with items, or with users. This part will contain an in-depth overview of different formulations and their natural applicability to different practical scenarios. In particular, we will focus on composite item formation in online shopping or recommendation applications, team formation scenarios in online collaborative applications, and group formation in social web analytics.

3.2 Optimization Algorithms

The second part of this tutorial (1h30mn) will describe theoretical results on the hardness of complex object formation and approximation and heuristic algorithms for both composite items and team/group. This part will also highlight the main experimental findings of proposed algorithms and will describe the symmetries between the two areas.

3.3 Future Directions

The third part of the tutorial (1hr) will describe new scenarios that combine both types of complex object formation (on items and on users) and the opportunities they give rise to in CC applications. Those examples are emerging in the web arena in particular in areas such as crowdsourcing complex tasks to teams, community feedback solicitation for complex items, and group discounts. We will review the challenges raised by the need to compose both items and users on the fly and the incremental maintenance of composed objects in an evolving scenario where new items enter the system and users arrive or leave.

4. AUDIENCE AND RELEVANCE

The tutorial will be of interest to both theoreticians and practitioners who are interested in the development of novel data-centric applications ranging from large-scale analytics to emerging complex crowdsourcing applications.

The proposed tutorial is timely as it addresses unsolved questions in the emerging area of complex crowdsourcing. 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 and Analysis, Graph Databases, Information Retrieval, and Modeling, Mining and Querying User Generated Content.

5. AUTHOR BIOGRAPHY

Sihem Amer-Yahia is DR1 CNRS at LIG in Grenoble where she leads the SLIDE team. Her interests are at the intersection of large-scale data management and data analytics. Earlier, she was Principal Scientist at the QCRI and 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., Editor-in-Chief of the VLDB Journal for Europe and Africa and is on the editorial boards of TODS and the Information System Journal. She is currently serving as PC chair of BDA 2015 and of SIGMOD Industrial 2015.

Senjuti Basu Roy is an Assistant Professor at the Institute of Technology at the University of Washington Tacoma. Her interests are data management, analysis, and algorithms. Prior to joining UW in January 2012, she was a postdoctoral fellow at DIMACS at Rutgers University. Her past industrial experience includes working at Microsoft Research and IBM Research. She has co-chaired KDD Cup 2013 and serves as a guest editor of JMLR special issue on KDD Cup 2013. She regularly serves on the program committees of the leading database and IR conferences and journals.