Computer Networks 90 (2015) 5–13

Contents lists available at ScienceDirect

Computer Networks

journal homepage: www.elsevier.com/locate/comnet

Hybrid human–machine information systems: Challenges and opportunities

Gianluca Demartini*

Information School, University of Sheffield, UK

ARTICLE INFO

Article history:
Received 31 October 2014
Revised 26 February 2015
Accepted 15 May 2015
Available online 8 July 2015

Keywords:
Human computation
Crowdsourcing
Database
Semantic web
Information retrieval

ABSTRACT

Micro-task Crowdsourcing has been used for different purposes: creating training data for machine learning algorithms, relevance judgments for evaluation of information systems, sentiment analysis, language translation, etc. In this paper we focus on the use of crowdsourcing as core component of data-driven systems. The creation of hybrid human–machine systems is a highly promising direction as it allows leveraging both the scalability of machines over large amounts of data as well as keeping the quality of human intelligence in the loop to finally obtain both efficiency and effectiveness in data processing applications.

Such a hybrid approach is a great opportunity to develop systems that are more powerful than purely machine-based ones. For example, it is possible to build systems that can understand sarcasm in text at scale. However, when designing such systems it is critical to take into account a number of dimensions related to human behavior as humans become a component of the overall process.

In this paper, we overview existing hybrid human–machine systems presenting commonalities in the approaches taken by different research communities. We summarize the key challenges that one has to face in developing such systems as well the opportunities and the open research directions to make such approaches the best way to process data in the future.

1. Introduction

With the rapid growth of data available in enterprises and on the Web, the need for effective and efficient data processing systems gets stronger. Data is a key asset in business and it has become key to support decisions. While machine-based solutions for large-scale data processing exists, they are limited in the type of data processing tasks they can do. Examples of tasks where machine-based systems perform poorly include image understanding, detecting opinions or sarcasm in text, etc.

To alleviate these problems, hybrid human–machine systems leveraging human intelligence at scale in combination with machine-based algorithms have been proposed. These systems make use of crowdsourcing by asking data related questions to a crowd of human individuals available to answer them. Thanks to such a human intelligence component, this type of information systems can perform tasks which are otherwise not possible to accomplish. Machine-based preprocessing or post-processing enables scalability over large amounts of data (e.g., thanks to scale-out architecture like Map/Reduce [18]).

Data chunks with related questions which are sent to the crowd by the system are usually called Human Intelligence Tasks (HITs) as they require human intelligence to be completed. A variety of task types is commonly published on these crowdsourcing platforms varying from audio transcription to general population surveys (see [31] for a classification).

* Tel.: +39 349 5119466.
E-mail address: demartini@exascale.info, g.demartini@sheffield.ac.uk

http://dx.doi.org/10.1016/j.comnet.2015.05.018
1389-1286/© 2015 Published by Elsevier B.V.

Crowdsourcing is a very general term covering topics from innovation [50] to citizen science [35]. Popular crowdsourcing examples include Wikipedia, a free on-line encyclopedia that anyone on the Web can edit; GalaxyZoo, a platform where any user can annotate large amounts of scientific images obtained with telescopes or from experiments [35]; and Recaptcha, used originally to correct OCR errors in a large book digitalization project [64].

In this paper we focus specifically on systems that leverage *paid micro-tasks crowdsourcing*. Commercial platforms like Amazon MTurk [36] have been built to support the exchange of HITs between *requesters* who need tasks to be completed and *workers*, that is, members of the crowd, who are willing to complete tasks motivated by a financial incentive.

In this paper we describe hybrid human–machine systems that crowdsource many small tasks to a crowd of human workers who complete them in exchange of a small monetary reward. We describe the most popular hybrid systems, their characteristics, and the main challenges that need to be faced when building a system with a crowd component inside. Aspects to be dealt with include controlling latency, data quality, and crowd motivation. Finally, we present a set of research directions in the area of hybrid human–machine systems. These include long-term use of crowds, complex hybrid data pipelines, and crowdsourcing efficiency improvements.

The rest of this paper is structured as follows. In Section 2 we overview existing hybrid human–machine systems proposed and evaluated by different research communities including database, information retrieval, social networks, semantic web, and data-driven sciences such as biomedicine and astrophysics. In Section 3 we summarize the main challenges that these types of systems have to face when dealing both with large amounts of data as well as with human individuals performing tasks for the system. In Section 4 we describe different open research questions in the area of human computation and crowdsourcing that need to be addressed to improve efficiency and effectiveness of hybrid human–machine systems. Finally, Section 5 concludes the paper.

2. Existing human–machine systems

Because of the ability to effectively process data at scale, a number of hybrid human–machine systems have been recently proposed within different data-related research fields. In this section we provide an overview of such systems.

2.1. Early human computation systems

Early examples of systems that leverage human intelligence in combination of machine-based data processing mostly leveraged the fun incentive rather than the monetary one. Thus, by means of gamification, systems like the ESP game were designed [62]. In this system two human players have to agree on the words to use to tag a picture without the possibility to interact with each other. Tags over which an agreement is reached are collected and used to generate a large collection of tagged images that can be used, for example, to train supervised machine learning algorithms. An extension of the ESP game is Peekaboom: a game that asks players to detect and annotate specific objects within an image [63]. A very popular crowdsourcing application is Recaptcha [64], which generates captcha codes that human users have to type to get access to Web content and which contain scanned words (from books) that would be otherwise complex to identify by means of automated OCR approaches. Thus, by entering valid captcha codes, human intelligence helps to digitize large amounts of textual content otherwise only available on paper. Recaptcha is now being used also for other purposes such as transcribing house numbers within pictures.

2.2. Data processing

The first crowd-powered database was CrowdDB proposed in 2011 by [30]. This system leverages crowdsourcing to process query operators within more powerful SQL queries that can, for example, retrieve images for a motivational slide show. In this case the crowd is used to tag images on their motivational dimension which is a relatively simple task for humans but a very complex one for machine-based algorithms. After this first foundational work, a number of more specific database problems have been addressed by hybrid human–machine approaches.

One of these is *entity resolution*. That is, detecting that two instances in the database refer to the same real-world entity (e.g., IBM and International Business Machines). In this context, proposed hybrid human–machine systems combine automatic approaches that compute similarity between large number of entity label pairs and crowdsource some entity pairs for manual matching thus obtaining both scalable and accurate entity resolution. To obtain this result it is important to minimize the number of HITs to be crowdsourced by leveraging machine-based algorithm confidence scores to selectively crowdsource entity-pairs to be matched [20]. In [66] authors show how an hybrid human–machine approach performs better than both a purely machine based approach as well as reduces the amount of human work to be done as compared to fully manual resolution. They also show how presenting the task in the form of a table containing multiple entities to be resolved instead of single entity pairs reduces the latency of the crowd. Related to this, [68] observed how allowing workers not to answer a specific entity resolution task improves the overall accuracy of the system. Also focusing on entity resolution, [67] studies how to estimate the accuracy gain obtained by each additional crowdsourced task. This is done to select the HITs that maximize the expected accuracy.

Another database related problem is that of *skyline queries*. These are complex-to-process queries that aim at retrieving optimal results over multiple dimensions. For example, hotels that are best in terms of price and distance to the beach. In this example, some results will always be worse than others in terms of both dimensions and can be safely filtered out in the early stages of query processing.

An hybrid human–machine approach has been proposed for this type of queries as well. In [44] authors focus on selecting which data items to crowdsource to obtain maximum result quality for skyline queries while controlling the cost of paid crowdsourcing. In detail, while finding missing values for all the tuples in a database may be not cost-efficient, by computing the Pareto optimality, it is possible to select