

# PCMA: Patterns for Crowdsourced Media Analysis

PATTERNS 2017, The Ninth International Conferences on  
Pervasive Patterns and Applications (IARIA conference)

Alessandro Ortis

Università degli Studi di Catania  
Viale A. Doria 6, 95125 - Catania, Italy  
Email: ortis@dmi.unict.it

*Abstract*—Nowadays, the diffusion of social networks plays a crucial role in collecting information about people opinion, trends and behaviour. The proliferation of mobile devices and the diffusion of social media, have changed the communication paradigm of people that share multimedia data, by allowing new interaction models. In particular, the rise of social media has opened new opportunities to better understand various aspects of the people’s interactions, preferences and behaviours. From such user contributed data, it’s possible to study the interrelationship between users, and to model and predict user behaviour (e.g., preferences, search intent, purchase behaviour). Advanced techniques for translating large datasets built from crowdsourced data into clear actionable insights that create value for business and society can be developed. Through the most common social media platforms, one can collect huge amount of multimedia data related to particular places and periods. These data includes a multitude of content types, such as texts, images, videos, GPS position etc. Moreover, the monitoring of such data over time gives useful information for further insights. This allows, for instance, the monitoring of specific events or sites of interest by the analysis of the multimedia activity of the users. This kind of analysis can be exploited on a large number of practical fields, such as event monitoring, marketing, computer forensics. This special session focuses on novel concepts, ideas, models, and systems able to take full advantage of the huge amount of the publicly available data generated by users every day.

*Keywords*—Crowdsourcing; Multimedia; Social Networks.

## I. INTRODUCTION

The aim of the PCMA 2017 special session is to bring together researchers from several sub-fields of Data Analysis working on the analysis of huge amount of multimodal crowdsourced data. With the rapid growth in communication technology, both companies and research institutes have been given the opportunity to perform large scale analysis on a multitude of real user-generated data, with a huge variety of application contexts. Crowdsourcing provides the opportunity for input from a number of sources, with different degrees of granularity. It allows organizations to develop solutions for both strategic issues and a method to find new ways to reach audiences on a broader scale. Moreover, the growing industry of online communication through smartphones provides a way for people of all backgrounds to give input on a project or research. Social media, blogs, forums, comment sections in online websites allow the opportunity for people to give suggestions or concerns.

There are three main assets that supported the rise of the “crowdsourcing era”:

- **Social Platforms:** the diffusion of social networks plays a crucial role in collecting information about people opinion, trends and behaviour. There are general social networks like Facebook where people chat, read news, and share their experiences. Furthermore, there are also very specific social platforms aimed to bring together people with common interests. There are platforms by which computer engineers share code and advices, or professional photographers can share their photos, etc. What happens now is that people love sharing their information, tell friends what they are doing and how they feel. And what is very important for the scientific community is that most of these information are public and immediately available.
- **Internet:** the number of people with an Internet connection is increasing, as well as the bandwidth and the available connection speed. With the 5G connection, it’s possible to download an high quality two hour long movie in less than 4 seconds. The connectivity improvements allowed the development of new services based on the transmission of huge amount of data, and real-time services. This allowed, for instance, web-based services like Netflix and the IP television, with the possibility to watch movies or live events with very high quality and low latency, or to perform a video of the event the user is attending allowing him to share the live streaming through a social network.
- **Personal Devices:** the diffusion of personal devices like smartphones allows people to be connected in every second of their lives, wherever they are in that moment. This allows the users to access on-line services in any moment of their daytime. Moreover, the amount of personal data that can be acquired by personal devices allow these services to be more pervasive and user centric.

Companies have been attempting innovative ways to get their customers involved both in production and promotion processes of their products and services. Crowdsourcing brings people together through a web-based platform, generally by means of social media, so businesses can obtain insights about what topics consumers are talking about or are interested in. Asking what people like before offering a new product on the market helps reduce the risk of a product or service failure, while also generating hype around a new offering.

In the last decade, several companies exploited the crowdsourcing paradigm to offer innovative services. For example, crowdsourcing has changed the way people travel. The rise of services like AirBnB, Uber, and what has been termed the “sharing economy”, transformed what had been primarily a mass-produced experience into a peer-to-peer economic network.

Companies like AirBnB and Uber have driven down prices by increasing the marketplace offer. Customers also benefit from increased variety and personalization in their travel options. The travellers issues and habits has remained rather the same, what have changed are only the service providers, and often times the service provided. Although the low prices can be attractive, the most of users trust the deals of such kind of companies due to the feedbacks of previous customers. Indeed, they do not actually trust the companies, but the opinions of other users of the community (preferably a large amount of them, specially if they are expert users of the platform who provided useful and fair feedbacks). On the other hand, these companies push users to public comments, express their opinions and tell their experiences by exploiting the “gamification” approach: the more you contribute, the more you earn (in terms of discounts, reputation, platform tools).

Besides new emerging companies, also the main IT companies have sought out innovative ideas to exploit crowdsourcing. Google exploits its users’ contributions to improve the quality of Google Translate results, and the GPS locations transmitted by a large number of users’ smartphones to infer traffic conditions in real time on major roads and highways.

In 2008, Facebook has exploited crowdsourcing to create different language versions of its website. The company claimed this method offers the advantage of providing site versions that are more compatible with local cultures. CitySourced [1] is an enterprise civic engagement platform. Citizens can identify and report civic issues, such as public works, quality of life, and environmental issues, and submit a report to the municipality through a mobile application. AstroDrone: is a scientific crowdsourcing project of the European Space Agency [2]. People who possess a Parrot AR drone can play the game, in which they are challenged to perform different space missions in an augmented reality environment. The analysis of the visual features extracted from the robot’s camera images is meant to improve robotic perception for tasks such as obstacle avoidance.

## II. SUBMISSIONS

The first paper presents a demo entitled “The Social Picture”. This framework allows to collect and explore huge amount of crowdsourced social images related to specific events or locations such as public events, cultural heritage sites and other customized private events. The Social Picture aims to create social communities of users that contribute to the creation of image collections about common interests, events and places. Computer Vision and Machine Learning algorithms are applied to perform Data Mining on such picture collections, and to organize them in a semantic way, to provide an easy exploration of large sets of images. The Social Picture collects huge number of public pictures obtained just by selecting a few inputs such as a location, radius, tags, and the period of crawling. Although the number of images could be huge, the system provides tools for the summary of the useful collection

insights and statistics. It is able to automatically organize the pictures in semantic groups, according to several and live customizable criteria. The Social Picture can be used as a tool for analysing the multimedia activity of the audience of an organized event, or the activity of people visiting a cultural heritage site, performing inferences on the attitude of the participating people. The obtained information can be then exploited by the event organizers for the event evaluation and further planning or marketing strategies.

Beside its aims and the number of provided tools, this framework provide a functional system to collect public image datasets, which can be automatically labelled by the system inference algorithms.

The second paper, presented by Marco Cavallo, introduces a work in progress project which aims to build a framework able to provide a quickly monitoring of critical events, obtained by exploiting a crowdsourcing approach in the context of Smart Cities. The proposed system defines a process that allows to collect, analyse and make specific inferences on crowdsourced data, with the aim to provide useful insights to the municipality.

The contributes of the users are filtered considering several factors including user distance with respect to the location of the monitored event, and the record of previous contributions of the users. In this sense, the user profiling is exploited with the aim to proper select a group of virtuous citizens, who have been effectively helped the community in the case of critical mass events. To better filter the amount and variety of contributes, the system proposes a fast survey to the users nearby the event. According to the acquired answers, the position of the users and their history, the systems selects a subset of contributors and asks them to send proper data, such as a live stream video of the event. The proposed system is fully integrated in a realistic Smart City scenario, where both dated and emerging technologies permit the realization of smart and efficient public services.

## III. CONCLUSION

The PCMA special session includes a broad range of topics related to the analysis of user generated data. Albeit they have not brought new technologies, the presented contributions have shown how existing technologies and resources can be efficiently combined with the crowdsourcing approach to provide inferences about audiences, events, locations, providing new interesting services. The presented works, proposed innovative systems able to distil knowledge and insights from large amount of crowdsourced data, providing interesting ideas for future work in this research domain.

## ACKNOWLEDGMENT

I would like to thank the organisers of PATTERNS 2017 for their tireless efforts and for accepting PCMA as a special session. I also thank the members of the programme committee for their work with the reviews and feedback. Last, but not least, I am very thankful to the authors for their very interesting contributions.

## REFERENCES

- [1] “CitySourced,” <http://www.citysourced.com/>, accessed: 2017-02-11.
- [2] “AstroDrone,” <http://www.esa.int/gsp/ACT/ai/projects/astrodrone.html>, accessed: 2017-02-11.