Special Issue on Big Personal Data in Interactive Intelligent Systems

F. CENA, Dept. of Computer Science, University of Torino, Italy
C. GENA, Dept. of Computer Science, University of Torino, Italy
G. HOUBEN, Dept. Software Technology, Delft University of Technology,
M. STROHMAIER, Dept. of Computer Science, University of Koblenz-Landau, Germany

This special issue on Big Personal Data in Interactive Intelligent Systems publishes selected work on Interactive Intelligent System research in the domain of Personal Big Data, i.e., vast amount of data regarding user personal activities, as for instance lifelogging or quantified self data.

This issue contains three contributions presenting not only technical solutions but also discussing explicitly the consequences of these solutions for users of interactive intelligent system, both from design and engineering point of view.

CCS Concepts: • **Information systems → Data management systems**; **Information storage systems; Information systems applications; World Wide Web; Information Retrieval**

Additional Key Words and Phrases: Big Personal Data, Small Data, Intelligent Systems, Intelligent User Interfaces

ACM Reference format:

.

DOI:

1 INTRODUCTION

This special issue on Big Personal Data in Interactive Intelligent Systems aims to publish the best current work on the Interactive Intelligent System research in the domain of on Personal Big Data, as for instance lifelogging and quantified self data. Increasingly vast amounts of data about people's interaction with the social and physical world are generated when people use social media, personal tracking devices, and the Internet of things. In this special issue we seek answers to questions such as: How can big personal data be collected, analyzed, and exploited so as to provide new or improved forms of interaction with intelligent systems – and what new issues have to be taken into account?

The question of how to process big personal data is challenging because of its sheer amount, its heterogeneous and possibly contradictory nature, and the semantic distance between the data and the world they represent, as well as the conclusions that can be drawn from it.

Big personal data can be put to good use in the service of users in various novel ways, but the question of what goals to pursue with it and how to pursue them is open-ended and not easy to answer.

Exploiting big personal data in interactive systems raises many issues of usability and acceptance, ranging from privacy issues to those of system comprehensibility and controllability.

The following topic dimensions indicate the range of work that have been considered relevant to the special issue. In particular, we individuate some classes of issues to be considered in relation to big personal data. For each issue, we present how the special issue's papers address them.

1) SOURCE OF DATA: What is the origin of the big personal data considered in this work? There are several sources of data about users that can contribute to big personal data:

- Users' behavior on social media sites or recommender systems in general [1]
- Users' traces (e.g., comments) and micro-traces (e.g., likes) left on the web [3]
- Users' interaction with objects that are part of the internet of things or use of wearable monitoring devices [2]

2) BENEFITS FOR USERS: What benefits of big personal data processing for end users are considered? The advantages for end users are related to the advanced services exploiting such data, such as:

- Support for choices about future behavior
- New forms of personalization and recommendation [1]
- Data aggregation from heterogeneous sources [3]
- Predicting future user trajectories from Big Personal Data (e.g. career, health, or activity prognosis)
- Novel services enabled by gathered data [2], for instance regarding smart cities, smart buildings, intelligent transports, etc.

3) ACCEPTABILITY: What issues of usability and acceptance are considered? Personal data is a very sensitive topic, since it is dealing with different facets of user life and private information. Thus, there is the issue to be accepted by user, in particular paying particular attention to**:**

- How to ensure adequate user understanding of processing of personal data
- How to ensure adequate privacy protection [3]
- How to ensure adequate predictability and comprehensibility of an interactive system's behavior
- How to visualize big personal data for end users [2]
- What novel input methods for collecting big personal data are promising

4) USER-RELATED TECHNICAL CHALLENGES: What user-related technical challenges are addressed? Technical challenge are related to:

- How to combine both historical and streaming data [1]
- How to personalize and re-use the collection and storage of personal data [3]
- Big Personal Data integration and sensor fusion [2]

5) PROCESSING TECHNIQUES: What processing techniques are applied to the big personal data? Big personal data contains a wide and heterogeneous range of data and is able to serve many purposes. The scientific implication is that since it is so complex, it requires novel modalities for processing it, including (but not limited to) adaptations of work from the following areas

- Data mining
- Machine learning [2; 3]
- User modeling and adaptation
- Recommendation
- Other forms of personalization [1]

6) SOCIAL IMPLICATIONS: To what extent are social aspects considered / analyzed, and how? There is a variety of social implications of personal big data gathering and processing, such as

- Relationship between individual and social behavior in interactive intelligent systems
- Analysis of Big Personal Data on different aggregation levels (individual vs. groups vs. populations)
- Social influence on individual behavior in interactive intelligent systems

We are hoping that the audience enjoys reading the articles contained in this special issue, and that it helps to inspire further research on personal big data.

2 LESSONS LEARNED FROM CONTRIBUTIONS

We received twelve contributions on different aspects related to big personal data (new services for users and citizens, improvement of traditional recommendation techniques), of which one of them [2] uses sensors as a source of data focusing on social web traces**.** We were able to accept three submissions for final publication [1; 2; 3] covering technical, design and HCI aspects to the above challenges. In the following we present the selected publications, highlighting how they tried to answer (parts of) these questions.

Zanzotto and Ferrone [3] tackle the problem of discovering threads in textual dialogues by introducing the novel task of discovering on-going conversations in scattered dialog blocks. Thus, they propose a publicly available testbed based on theatrical plays. They show that personal dialogs can be surrogated with theatrical plays - by trying to solve, in this way, the problem of privacy of Big Personal Data, since this testbed is sharable as it is based on public resources. They also propose a suite of computationally light learning models that can use syntactic and semantic features demonstrating that models for this challenging task should include features capturing shifts in language use and, possibly, modeling underlying scripts. Their models allowed to investigate a variety of features including coherence of linguistic registers and distributed and distributional representations for syntax and meaning, respectively.

With respect to the above classification the paper of Zanzotto and Ferrone [3] answers to our questions in this way:

1. SOURCE OF DATA: Web
2. BENEFITS FOR USERS: Discovering conversations in scattered dialog blocks
3. ACCEPTABILITY: Privacy protection
4. USER-RELATED TECHNICAL CHALLENGE: Re-use collection and storage of personal data
5. PROCESSING TECHNIQUES: Machine learning
6. SOCIAL IMPLICATIONS: None

The paper from Cassavia et al. [1] proposes intelligent solutions for analysing the huge amounts of data regarding user social interactions, user preferences and opinions. In particular they presents a user behavior oriented search framework tailored for analyzing user interactions with intelligent systems while seeking for some domain specific information (e.g., choosing a good restaurant in a visited area). The framework enhances a user's quest for information by exploiting previous knowledge about their social environment, the extent of influence the users are potentially subject to and the influence they may exert on other users. User influence spread, across the network, is dynamically computed as well to improve user search strategy by providing specific suggestions, represented as tailored faceted features. Such features are the result of data exchange activity (called data posting) that enriches information sources with additional background information and knowledge derived from experiences and behavioral properties of domain experts and users. The approach is tested in an application scenario of a tourist recommendation system, but it can be profitably exploited in several other contexts, e.g., viral marketing and food education.

With respect to the above classification the paper of Cassavia et al. [1] answers to our questions in this way:

1. SOURCE OF DATA: Recommender systems
2. BENEFITS FOR USERS: New forms of personalization
3. ACCEPTABILITY: None
4. USER-RELATED TECHNICAL CHALLENGES: How to combine both historical and streaming data
5. PROCESSING TECHNIQUES: Other forms of personalization
6. SOCIAL IMPLICATIONS: None

The focus of the work of Kim et al. [2] is on the provision of an intelligent interface to non-expert users for analyzing data coming from sensors. They presented a system implementing machine learning (ML) applications for sensor-based time series data as a novel domain-specific prototype that integrates interactive visual analytic features into the ML pipeline. Their tool is aimed at helping intermediate users to build systems involving classification of sensor data, greatly expanding the number and quality of these useful systems. They identify future directions for usable ML systems based on sensor data that will enable intermediate users to build systems that have been prohibitively difficult. Classification of sensor data remains a useful procedure that enables many important applications. However, this task remains difficult except for a small group of experts. Intermediate ML users still do not have adequate tools to help them build systems involving sensor data classification, due to the volume, multiple sources, and unintelligibility of sensor data. They evaluated the usefulness of multiple features and uncovered further recommendations for future interactive ML systems through introductory inquiries and two prototype based user studies,

With respect to the above classification the paper of Kim et al. [2] answers to our questions in this way:

1. SOURCE OF DATA: Sensors (wearable and internet of things)
2. BENEFITS FOR USERS: Novel services enabled by gathered data (for intermediate users)

3. ACCEPTABILITY: How to visualize big personal data
4. USER-RELATED TECHNICAL CHALLENGES: Sensor fusion (and classification)
5. PROCESSING TECHNIQUES Machine learning
6. SOCIAL IMPLICATIONS: None

ACKNOWLEDGMENTS

The guest editors would like to thank all the authors for their submissions. Furthermore, we would like to thank the reviewers who reviewed the submissions and helped to keep a high quality of the accepted papers. Last but not least, we want to thank, the past Editor-in-Chief of the journal, Anthony Jameson, for giving us the opportunity to edit this special issue and for helping us in all the stages of the work.

REFERENCES

[1] N. Cassavia, E. Masciari, C. Pulice, D. Sacca, Discovering User Behavioral Features to Enhance Information Search on Big Data, *this issue*., 2017

[2] S. Kim, D. Tasse, A. Dey, Making Machine Learning Applications for Time-Series Sensor Data Graphical and Interactive, *this issue*., 2017

[3] F. Zanzotto, L. Ferrone, Have you lost the thread? Discovering on-going conversations in scattered dialog blocks, *this issue*., 2017

Received February 2007; revised March 2009; accepted June 2009