



Research
iCity & Big Data—Perspective

Big Data Research in Italy: A Perspective

Sonia Bergamaschi^a, Emanuele Carlini^b, Michelangelo Ceci^c, Barbara Furletti^d,
Fosca Giannotti^d, Donato Malerba^{c,e,*}, Mario Mezzananza^f, Anna Monreale^d,
Gabriella Pasi^{g,*}, Dino Pedreschi^{d,h}, Raffele Perego^b, Salvatore Ruggieri^h

^a Department of Engineering “Enzo Ferrari,” University of Modena and Reggio Emilia, Modena 41125, Italy

^b High Performance Computing Laboratory, Institute of Information Science and Technologies of the Italian National Research Council (ISTI-CNR), Pisa 56124, Italy

^c Department of Computer Science, University of Bari Aldo Moro, Bari 70125, Italy

^d Knowledge Discovery and Data Mining Laboratory, ISTI-CNR, Pisa 56127, Italy

^e Big Data Laboratory, National Interuniversity Consortium for Informatics, Rome 00185, Italy

^f Department of Statistics and Quantitative Methods, University of Milano-Bicocca, Milan 20126, Italy

^g Department of Computer Science, Systems and Communications, University of Milano-Bicocca, Milan 20126, Italy

^h Department of Computer Science, University of Pisa, Pisa 56127, Italy

ARTICLE INFO

Article history:

Received 16 December 2015

Revised 4 June 2016

Accepted 13 June 2016

Available online 30 June 2016

Keywords:

Big data
Smart cities
Energy
Job offers
Privacy

ABSTRACT

The aim of this article is to synthetically describe the research projects that a selection of Italian universities is undertaking in the context of big data. Far from being exhaustive, this article has the objective of offering a sample of distinct applications that address the issue of managing huge amounts of data in Italy, collected in relation to diverse domains.

© 2016 THE AUTHORS. Published by Elsevier LTD on behalf of Chinese Academy of Engineering and Higher Education Press Limited Company. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

In the last few years, initiatives, events, and projects related to big data have proliferated, both in research centers/academies and in industry. The daily production of huge quantities of data related to various and diversified aspects of social life (including mobile phone data, social data, city related data, Web-based data, and health-related data) offers an unprecedented opportunity to observe and learn about peoples' preferences and behavior and to exploit this information in order to improve certain aspects of peoples' lives.

In response to this disruptive change—a change that opens up new economic perspectives—the European Commission has called

on national governments to wake up to the “big data” revolution[†]. The European digital economy has indeed been slow in embracing the data revolution compared with the US, and also lacks comparable industrial capability. To recover from this delay, considerable funding has been and will be provided by the European Commission as well as European countries to support research and innovation actions related to value generation from big data. To properly address this objective, various issues must be taken into consideration, ranging from the definition of powerful and technologically suited infrastructures to support-intensive data-driven computations (on both hardware and software) to the setting-up of multidisciplinary teams to properly and fruitfully extract knowledge from data in var-

* Corresponding authors.

E-mail addresses: donato.malerba@uniba.it; pasi@disco.unimib.it

[†] “Towards a thriving data-driven economy,” communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee, and the Committee of the Regions, Brussels, 2 Jul 2014.

ious domains.

Despite this delay, the European big data market controls the second largest market share, at 20% in terms of revenue in the global big data market [1]. Germany, the United Kingdom, France, and Italy are key countries in this market. In particular, the Italian big data market has grown rapidly in the last year, and significant investments are expected shortly from both the private and public sectors. This short survey reports some of the applications and projects that are undertaken by Italian universities with respect to the challenge of big data; in particular, it reports projects related to improving citizens' lives. As interesting examples of applications of technologies related to big data management, Section 4 describes a system aimed at monitoring the production and consumption of energy, while Section 5 synthetically presents a prototype aimed at analyzing Web job vacancies collected from five European Union (EU) countries and extracting the requested skills from the data.

It is important to note that this survey is far from being exhaustive, both with respect to active research groups on these topics (since such groups are considerably more numerous than those referred to in this paper) and with respect to projects, which in Italy are much more numerous than those reported here. The main aim is to offer the reader a flavor of the problems that academics are addressing with respect to this important issue.

A recent national initiative on big data is represented by the CINI "Big Data" Laboratory. CINI (www.consortio-cini.it) is the Italian National Interuniversity Consortium for Informatics, a consortium of 41 public Italian universities that promotes and coordinates scientific activities of research and technological transfer, both theoretical and applicative, in several fields of computer science and computer engineering. The consortium is a founding member of the Big Data Value Association (www.bdva.eu), the industry-led contractual counterpart to the European Commission for the outlining and implementation of the European strategic research agenda on big data. In addition, the CINI "Big Data" Laboratory—which focuses on data that is distributed throughout the whole national territory—has the aim of being a center of Italian expertise for the development of knowledge and technologies in the fields of big data and data science. Thirty-three Italian universities and about 300 researchers currently adhere to this initiative.

The next sections briefly present a few projects carried out by different Italian universities and research centers that address the issue of big data and that aim to improve various aspects of people's lives. These projects are related to distinct applicative domains, including understanding city dynamics, the Italian healthcare system, forecasting energy production in photovoltaic power plants, and managing job offers. The last two sections address the important issues of privacy and big data usability.

2. Understanding human and city dynamics with mobile phone data

Cities have always been complex systems of people, things, environments, and activities, and their rapid evolution has led to an unavoidable increase of complexity. This fact is pushing scientists to leave traditional paradigms of model-driven analysis in favor of data-driven approaches, opening the era of big data analytics. Digital signs that people produce every day by interacting with devices, social media, and other technological systems give unprecedented opportunities to study and understand city dynamics and social behavior from several perspectives. Understanding these dynamics means being able to anticipate the impact of phenomena and to support policies and planners in responding to citizens' needs.

Mobile phone data actually represent a proxy for studying and measuring cities and citizens, allowing us to identify peoples'

presence at the urban level [2–4], to reconstruct their mobility [5–8] and sociality [9], and to study the impact of events in cities [4,10].

2.1. Mobile phones and origin and destination (OD) matrix estimation

The estimation of presences and flows between preferred locations can be used to reconstruct an origin and destination (OD) matrix [5,6] that is useful for inferring transport demand models and for understanding infrastructure requirements. In Ref. [6], a long-term analysis of individual call traces is performed, in order to reconstruct systematic movements (i.e., movements with a high frequency) between the two most significant locations for an individual. Such locations, typically associated with home and work, are identified among the locations from which an individual made the largest number of calls. After having identified the systematic movements between these locations, the OD matrix summarizes the expected traffic flows between spatial regions.

2.2. Mobile phones for novel demography and city user estimation

The possibility of measuring and monitoring social phenomena has increased the interest in the use of big data to support official statistics [5]. Since administrative data cannot be collected with high frequency and often do not contain accurate information on mobility, calling data are being used more and more to integrate traditional data sources, allowing the construction, for example, of permanent observatories of the cities [3] and the identification of actual types of city users. In Refs. [2] and [4], the Sociometer, an analytical framework aimed at classifying mobile phone users into behavioral categories, is presented. The analytical process starts with the construction of spatio-temporal profiles synthesizing the presence of the individuals in the area of interest. Then, by applying a data mining method, different people categories are learned, and annotated profiles belonging to residents, dynamic residents, commuters, and visitors are produced. In Ref. [5], starting from the result of the Sociometer, an OD matrix at the municipality level is created in order to observe the inter-city mobility of the individuals. By producing statistics comparable to those obtained by the National Institute of Statistics (Italy), a safe way is offered to integrate existing population and flow statistics with the continuously up-to-date estimates obtained from mobile phone data.

2.3. Mobile phones, mobility diversity, and economic development

Studies become more and more challenging when there is a need to investigate society status in order to improve living conditions. In Ref. [8], starting from nation-wide mobile phone data, the authors extract a measure of mobility diversity and mobility volume for each individual, and investigate the correlations with external socioeconomic indicators. Diversity is defined in terms of the entropy of the individual users' trajectories, while volume of the mobility is measured by the characteristic distance traveled by an individual. The experiments show that mobility is correlated with wellbeing indicators (such as education level, unemployment rate, income, and deprivation), demonstrating a high predictive power of mobility behaviors with respect to the socioeconomic development of cities. In another exploration of the social dimension, an interesting result emerges in Ref. [9] from comparing mobility with the social network extracted from calls. The similarity in the movements and proximity in the social network appears to be strongly related, leading to the conclusion that people not connected in the network, but topologically close

Download English Version:

<https://daneshyari.com/en/article/478805>

Download Persian Version:

<https://daneshyari.com/article/478805>

[Daneshyari.com](https://daneshyari.com)