Chapter 6

Big Data and Privacy
State of the Art

Amine Rahmani
University of Algiers 1 Benyoucef Benkhadda, Algeria

ABSTRACT

The phenomenon of big data (massive data mining) refers to the exponential growth of the volume of data available on the web. This new concept has become widely used in recent years, enabling scalable, efficient, and fast access to data anytime, anywhere, helping the scientific community and companies identify the most subtle behaviors of users. However, big data has its share of the limits of ethical issues and risks that cannot be ignored. Indeed, new risks in terms of privacy are just beginning to be perceived. Sometimes simply annoying, these risks can be really harmful. In the medium term, the issue of privacy could become one of the biggest obstacles to the growth of big data solutions. It is in this context that a great deal of research is under way to enhance security and develop mechanisms for the protection of privacy of users. Although this area is still in its infancy, the list of possibilities continues to grow.

INTRODUCTION

The progress of new information technology and knowledge discovery allowing large scale pervasive surveillance over massive data sets had raised the need to such techniques permitting this technology to get more advanced without compromising privacy of users. Nowadays, privacy has been one of the most enduring issues associated with Big Data rising and digital electronic information spreading. There is a growing concern over information privacy. The world of privacy does
not rely only on avoiding observation or hiding personal matters and relationships, but it is more than that, it means the ability of sharing information selectively and not publicly. Every day, oceans of data are being collected over the planet using wireless sensors, and intelligent devices. These data are partly sensitive, some of the data are more sensitive than others. Other parts of the data can be considered as sensitive in some cases while not in other cases. People and researchers may have confused between anonymity and privacy. They may seem the same in context, but in reality they are quite different. One’s shopping habits are private information but not anonym while authorship of a political tract are anonym data but not private. However, privacy preserving domain is not limited only on sharing information without revealing sensitive knowledge about it, but even more, it expands to the protection of users’ rights, the ability to make intimate personal decisions without government interference is considered to be a privacy right, as is protection from discrimination on the basis of certain personal characteristics. This chapter offers a general overview about privacy preserving concepts and techniques focusing on ones used often in Big Data. The remainder of this chapter is addressed to the presentation and discussion of privacy in context, the debates of cybersecurity and privacy, and the presentation of general introduction of different techniques used to maintain privacy of users.

**Context of Privacy**

In the light of the advancement of Big Data, information technology had become a general threat to privacy. Privacy concerns had raised in the last few years opening new doors and challenges to scientific especially with the developing of such complex and advanced techniques of data mining. The promise of Big Data Analytics is, at first time, to offer information that can be used for good purposes for both individuals and societies. In fact, taking a closer look, we can realise that data mining had been developed in order to allow moving from information about individuals to generalizations that apply to broad classes. That means that data mining in practice should not pose any privacy risk. The real problem resides in the infrastructure used to support it. The more the data is complete and accurate the better are the results. Having complete, comprehensive and accurate data is what causes the raising of privacy issues. In other words, privacy issues rely on the misuse of existed data and not analytical algorithms. This section is addressed to the discussion of the context and outlines of the word “privacy”.

A standard definition of privacy in dictionaries is often “freedom from unauthorized intrusion”. However, according to the most of privacy laws, it is a concept that is only applied to “individually identifiable data”. Regarding to these two definitions, we can define privacy in two extremities. One is that good data is the one gives us

www.igi-global.com/e-resources/library-recommendation/?id=77

Related Content

Overview of PAKDD Competition 2007
www.igi-global.com/article/overview-pakdd-competition-2007/1801?camid=4v1a

Algebraic and Graphic Languages for OLAP Manipulations
www.igi-global.com/article/algebraic-graphic-languages-olap-manipulations/1798?camid=4v1a
Machine Learning in Studying the Organism’s Functional State of Clinically Healthy Individuals Depending on Their Immune Reactivity
Tatiana V. Sambukova (2013). *Diagnostic Test Approaches to Machine Learning and Commonsense Reasoning Systems* (pp. 221-248).
www.igi-global.com/chapter/machine-learning-studying-organism-functional/69411?camid=4v1a

Measuring the Financial Crisis in Local Governments through Data Mining
www.igi-global.com/chapter/measuring-financial-crisis-local-governments/44281?camid=4v1a