Chapter 57
Big Data for Prediction: Patent Analysis – Patenting Big Data for Prediction Analysis

Mirjana Pejic-Bach
https://orcid.org/0000-0003-3899-6707
University of Zagreb, Croatia

Jasmina Pivar
University of Zagreb, Croatia

Živko Krstić
Atomic Intelligence, Croatia

ABSTRACT

Technical field of big data for prediction lures the attention of different stakeholders. The reasons are related to the potentials of the big data, which allows for learning from past behavior, discovering patterns and values, and optimizing business processes based on new insights from large databases. However, in order to fully utilize the potentials of big data, its stakeholders need to understand the scope and volume of patenting related to big data usage for prediction. Therefore, this chapter aims to perform an analysis of patenting activities related to big data usage for prediction. This is done by (1) exploring the timeline and geographic distribution of patenting activities, (2) exploring the most active assignees of technical content of interest, (3) detecting the type of the protected technical according to the international patent classification system, and (4) performing text-mining analysis to discover the topics emerging most often in patents' abstracts.

INTRODUCTION

Patent databases are an abundant and important source of information about the particular technical field, and patent analysis has been proven as effective tool for decision makers who seek for a comprehensive overview of different technologies' topics, such as big data technologies (Madani & Weber, 2016). De-

DOI: 10.4018/978-1-6684-3662-2.ch057

cision makers may want to understand relevant trends, to spot new technologies in particular area or to estimate the importance of the emerging new technologies. Moreover, patent information is a relevant source for those who want to get familiar with key players of a particular technology, or to learn about their productivity and patenting behavior.

Big data technologies have attracted lots of attention due to their ability to analyze large amounts of various data sources, and extract useful information from them. Recently, big data technologies have become not only a methodology for analyzing the current situation, but are also used as tools for prediction in various fields, such as retailing, marketing and social media (e.g. Bradlow et al., 2017; Miah, Vu, Gammack & McGrath, 2017; Shirdastian et al., 2017).

Goal of this chapter is to analyze and help to understand patents related to big data for prediction. The paper will provide answers to the following questions that are of interest to big data inventors and investors: (1) What is the timeline of patents of big data solutions for prediction?; (2) Who are assignees of patents of big data solutions for prediction, and what is their geographic origin?; (3) What are the most frequent IPC patent areas of patents of big data solutions for prediction?, (4) What are the most often topics of patents of big data solutions for prediction? Answers to these questions will provide useful guidance related to competitiveness and new trends that emerge in the usage of big data technologies for prediction. Additional goal of this paper is to assess the usability of several data mining and text mining methods for the purpose of patent analysis, specifically association analysis of IPC patent areas, key-terms extraction and clustering. For this purpose, Statistica Text Miner 13.0, and Provalis Wordstat 8.0 has been used.

The chapter consists of the following sections. After the introduction, the second section presents the background of the research, encompassing the notion of big data, usage of big data for prediction, and usage of patent analysis. The third section describes the methodology used. The results of the analysis are presented in the fourth section. Finally, the last section is used to synthesise findings, present limitations, and future research directions of the chapter.

BACKGROUND

Big Data and Predictive Analytics

Big data has become an exciting field of study for practitioners and researchers, due to the need to adapt to the emergence of huge databases (Parr Rud, 2011). Each of them has different focus and concerns in this area, which yielded various definitions and descriptions of big data. Practitioners, such as consulting companies and multinational corporations, define big data by mainly focusing on the technology necessary to handle such data. For example, the National Institute of Standards and Technology describes it as data that exceed capacity or capability of conventional systems and "require a scalable architecture for efficient storage, manipulation and analysis" (NIST, 2017, p. 8). On the other hand, scientists describe big data as the phenomenon related to various characteristics of data generated by different actions, e.g. social media and business transactions. Boyd and Crawford (2012, p. 662) define big data as "cultural, technological, and scholarly phenomenon that rests on the interplay of technology, analysis and mythology". Furthermore, scientists often use following three characteristics in order to describe big data: Volume, Variety and Velocity. Volume describes the large amount of data that depends on the type of data, time and industry, which "make it impractical to define a specific threshold for big data volumes" (Gandomi