Expert Systems with Applications 42 (2015) 562–572

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Classification of countries' progress toward a knowledge economy based on machine learning classification techniques

Mónica de la Paz-Marín [a,*], Pedro Antonio Gutiérrez [b], César Hervás-Martínez [b]

[a] Department of Management and Quantitative Methods, Loyola Andalucía University, Business Administration Faculty (ETEA), Escritor Castilla Aguayo 4, 14004 Córdoba, Spain
[b] Department of Computing and Numerical Analysis, University of Córdoba, Albert Einstein Building, 14071 Córdoba, Spain

ARTICLE INFO

Article history:
Available online 20 August 2014

Keywords:
Decision support systems
Machine learning
Knowledge economy
Hierarchical clustering
Ordinal classification

ABSTRACT

Knowledge is a key factor of competitive advantages in the current economic crisis and uncertain environment. There are a number of indicators to measure knowledge advances, however, the benefits for stakeholders and policy makers are limited because of a lack of classification models. This paper introduces an approach to classify 54 countries (in 2007–2009) according to their progress toward a knowledge economy (KE). To achieve this, the aims of this paper are twofold: first, to find clusters of countries at a similar stage of development toward KE to test if they are meaningful; hence, it will be possible to order the clusters from early KEs (last cluster) to advanced KEs (first cluster). Second, having obtained these clusters, it is possible to build various models to detect the advancement of countries toward KE from one year to another due to its classification. Then, three ordinal classifiers from the machine-learning field were compared in order to select the classifier that performs the best and to confirm the ordinal description of the clusters. Finally, an ordinal model based on the Support Vector Ordinal Regression with Implicit Constraints was selected because of its ability to classify the patterns into the clusters, confirming the appropriateness of the clusters and their ordinal nature. The proposed ordinal classifier could be used for monitoring the progress or stage of transition to KE and for analysing whether a country changes clusters, entering one that performs better or worse.

1. Introduction

Academics, policy-makers, stakeholders, consultants and the media have shown a growing interest in the relevance of knowledge creation as a key factor enabling an increase in the competitive advantages of firms and, consequently, of national economies (Von Krogh, Ichijo, & Nonaka, 2000). That is especially crucial in the uncertain, changing, ambiguous and complex environment that characterises nations today (Johannessen & Olsen, 2010). Thus, it can be said that we are merging into the so-called knowledge economy or knowledge-based economy (KE), seen as the stage that follows the industrial era, which has become almost an imperative for nations, stressing even more the role of innovation in efforts to achieve competitiveness and a sustainable economic development. The fact that governing bodies place knowledge at the core of their strategies also reveals the relevance of achieving this type of growth model.

In the last two decades, literature and research related to KE have proliferated (Aghion & Howitt, 1992; David & Foray, 2002; Drucker, 1993; Grossman & Helpman, 1991; Leydesdorff, 2006; OECD, 1996; Thurow, 1999), focusing mainly on the important role of knowledge or human capital as a source of long-term economic growth. The relevance of knowledge is clearly linked to a new growth theory, which considers knowledge (or human capital) to be an endogenous variable of economic growth. Knowledge is regarded as the basic form of capital and economic growth is driven by the accumulation of knowledge (Lucas, 1988; Romer, 1990). Other economic theories appear to examine this phenomenon: the evolutionary theory of economic change (Nelson & Winter, 1982), the national innovation systems theory (Freeman, 1987; Nelson, 1993), the knowledge gap theory (Abramovitz, 1986), the triple helix theory (Etzkowitz & Leydesdorff, 2000; Leydesdorff, 2005) or even the *N*-Duple of helices theory (Leydesdorff, 2006).

There are several real examples of the influence of knowledge on current economic growth: (1) progress in information and communications technology (ICT) that enables cheap and rapid access to knowledge and information; (2) the ever-increasing speed of scientific and technological advances; (3) global competition; and

* Corresponding author. Tel.: +34 957 2182 12; fax: +34 957 218030.
E-mail addresses: mpaz@uco.es (M. de la Paz-Marín), pagutierrez@uco.es (P.A. Gutiérrez), chervas@uco.es (C. Hervás-Martínez).

http://dx.doi.org/10.1016/j.eswa.2014.08.008
0957-4174/

(4) the new demands, tastes and customs of citizens. Based on all these, the World Bank Institute emphasises that most countries that have made rapid progress staged nationwide KE-inspired change programs (International Bank for Reconstruction, 2007).

The concept of knowledge economy (KE), although it may have its roots in Adam Smith's work, was possibly first used by Machlup (1962) and coined by Drucker (1969) and is an object of special attention in the KE report of the Organisation for Economic Co-operation and Development (OECD). According to this report, knowledge economies are those 'which are directly based on the production, distribution and use of knowledge and information' (OECD, 1996, pp. 7). The definition of the World Bank is also a widely used setting that is 'essentially an economy where knowledge is the main engine of economic growth' (Chen Derek & Dahlman, 2006, pp. 1). In these economies, emphasis is placed on intellectual capabilities rather than physical factors (Powell & Snellman, 2004) or, similarly, the share of intangible capital is greater than that of tangible capital in the overall stock of real capital (Foray, 2004).

In this context, governments must plan investments and develop strong education systems to train highly skilled workers for highly skilled jobs if they seek to achieve a knowledgeable society (Hsu, Lin, & Wei, 2008). Measurement tools, frameworks, models and methodologies help stakeholders to analyse and benchmark the capabilities of countries as knowledge-based economies. Such assessments facilitate the adoption of policies as well as the creation of national knowledge systems for holistic development.

Numerous composite indicators have been created by many organisations including the World Economic Forum (WEF), the United Nation (UN), the World Bank (WB) or the International Institute for Management Development (IMD), to name a few. These indicators have been utilised by organisations including government agencies, aid agencies and research institutions to assess the competitiveness of a nation or nations in the context of KBE. However, these indicators suffer from many shortcomings, as they can be inconsistent as they generate different ranking and scores depending on the nature and type of assessments. Moreover, in all of these studies indicators have inherited two problems. The first is the definition of the weighting scheme and the second is that they are examined at a specific time (Mimis & Georgiadis, 2013).

These indicators yield different scores and rankings depending on the nature and type of assessments, report on past performance (Al Shami, Lotfi, & Coleman, 2012), which involves many questions that must be answered subjectively (Booysen, 2002) and do not anticipate the classification of the countries where a certain KE is heading (or could head) if all variables employed in the model were known.

Classification, in general, is one of the most frequent decision-making tasks in human activity. A (supervised) classification problem occurs when an object needs to be assigned into a class based on a number of observed attributes related to that object. The most common approach to classification considers that a class variable is composed of non-ordered labels, i.e., a variable exhibits a nominal nature and the categories cannot be ordered. However, many multi-criteria classification problems involve classifying data into classes that have a natural order (ordinal problem) (Zopounidis & Doumpos, 2002). Ordinal classification techniques have broad applications in which it is natural to rank instances such as information retrieval (Chu & Keerthi, 2007; Herbrich, Graepel, & Obermayer, 1999), econometric modelling (Mathieson, 1995), credit risk (Doumpos, Kosmidou, Baourakis, & Zopounidis, 2002; Xu, Zhou, & Wang, 2009) or gen analysis (Pyon & Li, 2009), to name a few.

Consequently, in this study, the problem has been addressed by using ordinal classifiers in order to test which ordinal model performs best. A priori, the dependent variable (the cluster or class previously obtained) has an ordinal consideration as can be seen in the myriad of examples that imply a ranking of countries in socioeconomic issues: the current Rating Agencies (i.e., Moody's, Standard and Poor's or Fitch), the Global Competitiveness Index of the International Monetary Fund, the Knowledge Economy Index of the World Bank, the Innovation Union Scoreboard of the European Commission, the Human Development Index of the United Nations, the ranking of universities, and so on.

In accordance with the above, the first aim of this work focuses on obtaining homogeneous groups of countries in relation to their progress toward KE. Thus, a hierarchical clustering (an unsupervised algorithm) was applied to detect behavioural patterns. As a result, a number of clusters were set and the characteristics of each one were defined. The second aim of this paper is to build a model for the classification of 162 country–year observations (54 countries in 2007, 2008 and 2009) thanks to their assignment into clusters previously obtained. In relation to the description of the clusters, it is clear they present an ordinal nature. For that reason, three ordinal classifiers were built to assign each country–year observation to its corresponding cluster and the results were compared to evaluate which performed the best due to the ordinal nature of this socioeconomic problem (see Fig. 1).

Ordinal classification algorithms yielded very good performance, and a Support Vector ordinal model was selected for the classification of countries into one of the clusters obtained. This model could help to monitor national strategies and some key features related to knowledge creation and innovation in general terms, analysing the evolution (or lack thereof) of the country to a better or worse cluster in terms of KE progress, because the cluster has an order similar to rankings.

The remainder of this paper is organised as follows: we briefly review the relevant literature on the assessment of knowledge economy in countries and classification methodologies in Section 2. Then, the methodology applied in this study is detailed in Section 3.

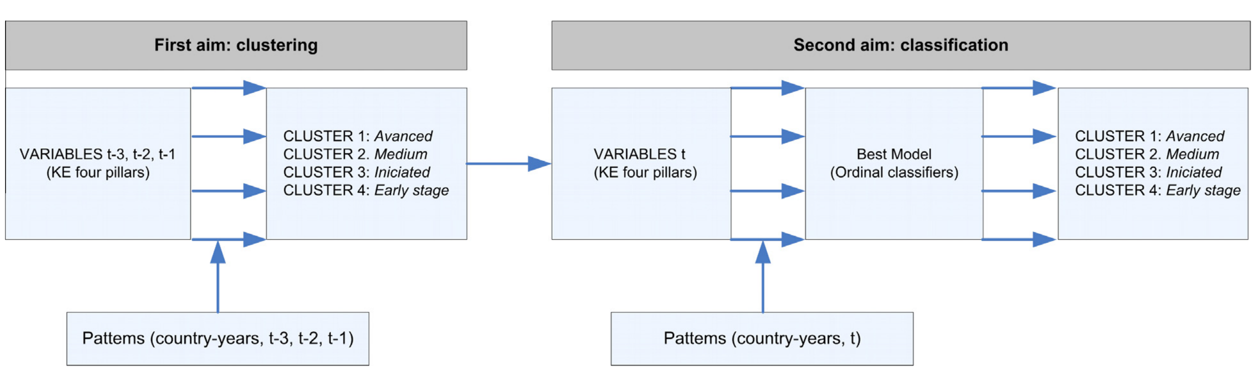

Fig. 1. Main stages and aims of the research.