A Cognitive Approach to Scientific Data Mining for Syndrome Discovery: A Case-Study in Dermatology

Francesco Gagliardi, University of Rome “La Sapienza,” Italy

ABSTRACT

The author introduces a machine learning system for cluster analysis to take on the problem of syndrome discovery in the clinical domain. A syndrome is a set of typical clinical features (a prototype) that appear together often enough to suggest they may represent a single, unknown, disease. The discovery of syndromes and relative taxonomy formation is therefore the critical early phase of the process of scientific discovery in the medical domain. The system proposed discovers syndromes following Eleanor Rosch’s prototype theory on how the human mind categorizes and forms taxonomies, and thereby to understand how humans perform these activities and to automate or assist the process of scientific discovery. The system implemented can be considered a scientific discovery support system as it can discover unknown syndromes to the advantage of subsequent clinical practices and research activities.

Keywords: Cluster Analysis, Cognitive Artificial Intelligence, Computational Scientific Discovery, Erythematous-Squamous Diseases, Prototype-Theory, Scientific Data Mining

1. INTRODUCTION

The process of scientific discovery has long been viewed as the pinnacle of creative thought. Thus to many people, including some scientists themselves, it seems an unlikely candidate for automation by computer. However, researchers in artificial intelligence have repeatedly questioned this attitude and attempted to develop intelligent artefacts that replicate the act of discovery.

The computational study of scientific discovery has taken important strides in its short history (Alai, 2004); the initially more influential approach was that of Herbert Simon and several co-workers (Langley et al., 1987). Their claim was that scientific discovery is a complex form of problem solving, and as such it can be simulated through computer programs in heuristic programming (see, for a critical analysis, Cordeschi, 1992; Gillies, 1996; Trautteur, 1992). A different approach is taken by Paul Thagard and others (Holland et al., 1986; Thagard, 1998a).

DOI: 10.4018/jssci.2012010101
who consider the process of knowledge-acquisition and discovery not primarily as a problem solving process, but rather as the result of creative reasoning (Holyoak & Thagard, 1995) which is based on analogical reasoning (Gentner et al., 2001; Holland et al., 1986, Ch. 10) and conceptual-consistency processes (Thagard, 2007).

Knowledge-acquisition and discovery-process concerns have also been raised on the one hand in machine learning and database technologies on data mining (Cios et al., 1998; Fayyad et al., 1996; Haibo et al., 2011) and on the other hand within the fields of cognitive informatics (Wang, 2007) and cognitive computing (Wang et al., 2010).

The techniques developed in these fields have been applied to the discovery of scientific knowledge, and are used in the fields of computational scientific discovery (Džeroski & Todorovski, 2007), scientific data mining (Gaber, 2009) and cognitive computing (Dartnell et al., 2008; Zhao et al., 2008); this research brings out interesting connections with the philosophy of science (Gagliardi, 2009; Korb, 2004; Williamson, 2009) and can have wide applications in the medical domain (Cios & Moore, 2001, 2002) and more recently in the field of cognitive computing applied to medical domain (do Espírito Santo et al., 2009; Anitha et al., 2010).

Early research focused on computational scientific discovery, replicating discoveries from the history of disciplines as diverse as mathematics, physics, chemistry and biology, as the collection by Shrager and Langley (1990) reveals. It therefore mainly attempted to model and replicate the historical record.

Researchers in the field of scientific data mining have focused their energies on the computational discovery of new scientific knowledge, and at the same time emphasized cooperation between intelligent artefacts and humans in this enterprise (Bridewell & Langley, 2010; Džeroski et al., 2007; Langley, 1998, 2000) based on an intriguing and synergetic interaction between natural intelligence and cognitive informatics (Dartnell et al., 2008; Zhao et al., 2008).

On the whole, these research fields seek to understand the products and processes of science by studying artefacts that engage and assist in knowledge construction. Along these lines, researchers have investigated activities as taxonomy formation, law discovery, and theory development. Their findings have demystified these activities and suggest a strong link between the disciplined practices of scientists and the everyday reasoning skills shared by everyone (Bridewell & Langley 2010, p. 36).

There are two primary reasons why we might want to study scientific discovery from a computational perspective (Džeroski et al., 2007, p. 2):

- To understand how humans perform this intriguing activity, which belongs in the realm of cognitive modelling; and
- To automate or assist in facets of the scientific process, which belongs in the realm of artificial intelligence.

So this approach can be soundly posed in the field of cognitive informatics (Wang, 2007), which is the field that investigates the internal information processing mechanisms of the natural intelligence, as well as their engineering applications in cognitive computing (Wang et al., 2010).

In this our study we introduce a machine learning system to discover syndromes which at the same time embed cognitive theory about how the human mind classifies and forms taxonomies and so we adopt this interdisciplinary view based on an interlacing between artificial intelligence and cognitive science.

This methodology is well-rooted in the ‘synthetic method’ (Boden, 2006; Cordeschi, 2002; Cordeschi, 2008; Gagliardi, 2007), i.e., the method of model building to understand intelligent behaviour, which starts from precursory ideas of the scientist as Craik (1943) or Hull with his
An Improved Particle Swarm Optimization Algorithm Based on Quotient Space Theory
www.igi-global.com/article/improved-particle-swarm-optimization-algorithm/72877?camid=4v1a

Analyzing Animated Movie Contents for Automatic Video Indexing
www.igi-global.com/chapter/analyzing-animated-movie-contents-automatic/49111?camid=4v1a