Bisociative Knowledge Discovery

Michael R. Berthold

Nycomed Chair for Bioinformatics and Information Mining,
Dept. of Computer and Information Science,
University of Konstanz, Konstanz, Germany
Michael.Berthold@Uni-Konstanz.DE

Abstract. Data analysis generally focusses on finding patterns within a reasonably well connected domain of interest. In this article we focus on the discovery of new connections between domains (so called *bisociations*), supporting the creative discovery process in a novel way. We motivate this approach, show the difference to classical data analysis and conclude by briefly illustrating some types of domain-crossing connections along with illustrative examples.

1 Motivation

Modern data analysis enables users to discover complex patterns of various types in large information repositories. Together with some of the data mining schema, such as CRISP-DM and SEMMA, the user participates in a cycle of data preparation, model selection, training, and knowledge inspection. Many variations on this theme have emerged in the past, such as Explorative Data Mining, Visual Analytics, and many others but the underlying assumption has always been that the data the methods are applied to models one (often rather complex) domain. Note that by *domain* we do not want to indicate a single feature space (Multi View Learning or Parallel Universes are just two of many other types of learning methods to operate on several spaces at the same time) but instead we want to emphasize the fact that the data to be analyzed represents objects that are all regarded as representing properties under one more or less specific aspect.

However, methods that support the discovery of connections between previously unconnected (or only loosely coupled) domains have not received much attention in the past. However, in order to really support the discovery of novel insights finding connections between previously unconnected domains promises true potential. Research on (computational) creativity strongly suggests that this type of "out of the box thinking" is an important part of the human ability to achieve truly creative discoveries.

In this paper we summarize some more recent work focusing on the discovery of such domain-crossing connections. To contrast the finding of "within domain" patterns (also termed associations) we use the term *bisociation* as coined by Arthur Koestler in [4] to stress the difference. We argue that *Bisociative Knowledge Discovery* represents an important challenge in our quest to building truly creative discovery support systems.

J. Gama, E. Bradley, and J. Hollmén (Eds.): IDA 2011, LNCS 7014, pp. 1–7, 2011.

2 Bisociation

Defining bisociation formally is, of course, a challenge. An extensive overview of related work, links to computational creativity and related areas in AI as well as a more thorough formalization can be found in [3]. Here we will only concentrate on the essential parts for the remainder of this paper and only intuitively motivate the background.

Boden [2] distinguishes three different types of creative discoveries: Combinatorial, Exploratory, and Transformational Creativity. Where the second and third category can be mapped on (explorative) data analysis or at least the discovery process within a given domain, Combinatorial Creativity nicely represents what we are interested in here: the combination of different domains and the creative discovery stemming from new connections between those domains.

Informally, bisociation can be defined as (sets of) concepts that are bridging two otherwise not –or only very sparsely– connected domains whereas an association bridges concepts within a given domain. Of course, not all bisociation candidates are equally interesting and in analogy to how Boden assesses the interestingness of a creative idea as being new, surprising, and valuable [2], a similar measure for interestingness can be specified when the underlying set of domains and their concepts are known. Going back to Koestler we can summarize this setup nicely:

> The creative act is not an act of creation in the sense of the Old Testament. It does not create something out of nothing; it uncovers, selects, re-shuffles, combines, synthesizes already existing facts, ideas, faculties, skills. The more familiar the parts, the more striking the new whole.

Transferred to the data analysis scenario, this puts the emphasis on finding patterns across domains whereas finding patterns in the individual domains themselves is a problem that has been tackled already for quite some time. Put differently, he distinguishes associations that work within a given domain (called *matrix* by Koestler) and are limited to repetiveness (here: finding other/new occurrences of already identified patterns) and bisociations finding novel connections crossing independent matrices (domains).

3 Types of Bisociation

Obviously the above still remains relatively vague and for concrete implementations the type of bisociative pattern that are sought needs to be specified better. In the past years a number of bisociation types emerged in the context of Bisociative Knowledge Discovery: Bridging Concepts, Bridging Graphs, and Bridging by Structural Similarity. Since these ideas are also addressed in other areas of research, additional types most likely exist in those fields as well.