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.

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 repetitiveness (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.

3.1 Bridging Concepts

The most natural type of bisociation is represented by one concept which links two domains, Figure 1 illustrates this.

Fig. 1. Bridging concept (from [3])

Such bridging concepts do not need to exist in the context of a network based representation as suggested by the figure but can also be found in other representations. In [6], for instance, different textual domains were analyzed to find bisociative terms that link different concepts from the two domains.

An example for a few bridging concepts is shown in Figure 2. Here a well known data set containing articles from two domains (migraine and magnesium) was searched for bridging terms. This example does reproduce an actual discovery in medicine.

Fig. 2. Bridging concepts - an example reproducing the Swanson discovery (from [6])

3.2 Bridging Graphs

More complex bisociations can be modeled by bridging graphs, Figure 3a illustrates this.

Fig. 3. Bridging graphs (from [3])

Here two different domains are connected by a (usually small) subset of concepts that have some relationship among themselves. In a network based representation one would identify a relatively dense subgraph connecting two domains but also in other representations such “chains of evidence” can be formalized, connecting two domains.

Two examples for bridging graphs are shown in Figure 4 (the data stems from Schools-Wikipedia, see [7] for details). One can nicely see how the two concepts “probability space” and “arithmetic mean” connect the domain of movies with some more detailed concepts in the statistics domain. This is at first glance surprising but finds its explanation in the (in both cases also somewhat “creative”) use of those concepts in the two films or the series of films dominated by one actor. The second example nicely bridges physical properties and usage scenarios of phonographs.

Fig. 4. Bridging graphs - two examples (from [5])

3.3 Bridging by Structural Similarity

The third, most complex type of bisociation does not rely on some straightforward type of link connecting two domains but models such connections on a higher level. In both domains two subsets of concepts can be identified that share a structural similarity. Figure 5 illustrates this – again in a network based representation but also here other types of structural similarity can be defined.

Fig. 5. Bridging by graph similarity (from [3])

An interesting example of such structural similarities can be seen in Figure 6. Again, the demonstration data set based on Schools-Wikipedia was used. The two nodes slightly off center (“Euclid” on the left and “Plato” on the right) are farther apart in the original network but share structural properties such as being closely connected to the hub of a subnetwork (“mathematics” vs. “philosophy”). Note that also “Aristotle” fills a similar role in the philosophy domain.

Fig. 6. Bridging by graph similarity - example (from [7])

3.4 Other Types of Bisocation

The bisociation types discussed above are obviously not complete. The first two types are limited to a 1:1 match on the underlying structures and require the two domains to already have some type of (although sparse) neighborhood relation. Only the third type allows matching on a more abstract level, finding areas of structural similarity and drawing connections between those. Other such, more abstract types of bisociation certainly exist but likely also more direct bisociation types can be defined as well.

4 Bisociation Discovery Methods

In order to formalize the types of bisociations and develop methods for finding them, a more detailed model of the knowledge space needs to be available. When dealing with various types of information and the ability to find patterns in those information repositories a network based model is often an appropriate choice due to its inherent flexibility. A number of methods can be found in [1]. We hasten to add, however, that this is not the only way to model domains and bisociations, again in [1] some contributions finding bisociation in non-network type domains can be found.

It is interesting to note that quite a few of the existing methods in the machine learning and data analysis areas can be used with often only minor modifications. For instance, methods for item set mining can be applied to the detection of concept graphs and measures of bisociation strength can also be derived from other approaches to model interestingness. Quite a bit of Bisociative Knowledge Discovery can rely on existing methods but the way those methods are applied is often radically different. Instead of searching for patterns that have reasonably high occurrence frequencies we are often interested in the exact opposite: bisociations are at their heart something that is new and only exists in very faint ways if at all so far.

5 Outlook

Bisociative Knowledge Discovery promises great impact especially in those areas of scientific research where data gathering still outpaces model understanding. Once the mechanisms are well understood the task of data analysis tends to change and the focus lies much stronger on (statistically) significant and validated patterns. However, in the early phase of research, usually the ability to collect data by far outperforms the experts ability to make sense out of those gigantic data repositories and use them to form new hypotheses. Current methods fall short of offering true, explorative access to patterns within but in particular across domains – the framework sketched here (and more substantially founded in [1]) can help to address this shortcoming. Much work still needs to be done, however, as many more types of bisociations can be formalized and many of the

existing methods in the Machine Learning and Data Analysis/Mining community are waiting to be applied to these problems.

One very interesting development here are the network based bisociation discovery methods which nicely begin to bridge the gap between solidly understood graph theoretical algorithms and overly heuristic, poorly controllable methods. Putting those together can lead to the discovery of better understood bisociative (and other) patterns in large networks.

The Data Mining Community has been looking for an exciting “Grand Challenge” for a number of years now. Bisociative Knowledge Discovery could offer just that: inventing methods and building systems that support the discovery of truly new knowledge across different domains will have immense impact on how research in many fields can be computer supported in the future.

Acknowledgements. The thoughts presented in this paper would not have emerged without countless, constructive and very fruitful discussions with the members of the BISON Project and the BISON Group at Konstanz University. In particular, I want to thank Tobias Kötter, Kilian Thiel, Uwe Nagel, Ulrik Brandes, and our frequent guest bison Werner Dubitzky for many discussions around the nature of bisociation and creative processes. Most of this work was funded by the European Commission in the 7th Framework Programme (FP7-ICT-2007-C FET-Open, contract no. BISON-211898).

References

1. Berthold, M.R. (ed.): Bisociative Knowledge Discovery, 1st edn. LNCS. Springer, Heidelberg (in preparation)
2. Boden, M.A.: Précis of the creative mind: Myths and mechanisms. *Behavioural and Brain Sciences* 17, 519–570 (1994)
3. Dubitzky, W., Kötter, T., Schmidt, O., Berthold, M.R.: Towards creative information exploration based on Koestler’s concept of bisociation. In: Berthold, M.R. (ed.) *Bisociative Knowledge Discovery*, 1st edn. LNCS. Springer, Heidelberg (in preparation)
4. Koestler, A.: *The Act of Creation*. Macmillan, NYC (1964)
5. Nagel, U., Thiel, K., Kötter, T., Piatek, D., Berthold, M.R.: Bisociative discovery of interesting relations between domains. In: *Proceedings of IDA the 10th Conference on Intelligent Data Analysis*, Portugal, Porto (in press)
6. Sluban, B., Juršič, M., Cestnik, B., Lavrač, N.: Exploring the power of outliers for cross-domain literature mining. In: Berthold, M.R. (ed.) *Bisociative Knowledge Discovery*. LNCS. Springer, Heidelberg (in preparation)
7. Thiel, K., Berthold, M.R.: Node similarities from spreading activation. In: Berthold, M.R. (ed.) *Bisociative Knowledge Discovery*, 1st edn. LNCS. Springer, Heidelberg (in preparation)