Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI-16)

A Visual Semantic Framework for Innovation Analytics

Walid Shalaby
Computer Science Department
University of North Carolina at Charlotte
wshalaby@uncc.edu

Kripa Rajshekhar
Metonymy Labs
Chicago Metropolitan Area
kripa@metolabs.com

Wlodek Zadrozny
Computer Science Department
University of North Carolina at Charlotte
wzadrozn@uncc.edu

Abstract

In this demo we present a semantic framework for innovation and patent analytics powered by Mined Semantic Analysis (*MSA*). Our framework provides cognitive assistance to its users through a Web-based visual and interactive interface. First, we describe building a conceptual knowledge graph by mining user-generated encyclopedic textual corpus for semantic associations. Then, we demonstrate applying the acquired knowledge to support many cognition and knowledge based use cases for innovation analysis including technology exploration and landscaping, competitive analysis, literature and prior art search and others.

Introduction

Innovations are often expressed as patent applications which tend to be lengthy documents with highly complex and domain specific terminology. To establish their work novelty, patent authors tend to use different vocabulary to refer to the same concepts making the problem of patent analysis linguistically challenging. Moreover, innovations analysis, management, and planning are cognition intensive tasks that require huge amount of domain and real-world expertise.

With the emergence of Big Data, massive amounts of innovation and patent data have become readily available promoting the need for more intelligent, interactive, and insightful cognitive analytics.

Current tools for innovation analysis can be roughly subdivided into three categories. First, statistical analysis tools which utilize innovations metadata to provide high-level statistics and visual plots of innovation trends. Second, content analysis tools which analyze patent texts to support tasks like prior art search, litigation, technology exploration and others. Third, hybrid tools which serve both purposes.

In this demo we present a Web-based semantic, visual, and interactive framework for innovation analytics. To address the linguistic challenges of patent texts, our framework utilizes Minded Semantic Analysis (*MSA*), a novel distributional semantics approach (under review). *MSA* builds a conceptual knowledge graph using association rule mining on concept rich user-generated textual corpora (e.g., *Wikipedia*). In this knowledge graph nodes represent concepts and edges represent associations between them. For any input keyword(s)/text, our framework maps that input to a conceptual domain using the knowledge graph. In this conceptual domain semantically similar terms/documents will have similar conceptual representations.

The basic pipeline for building the concept graph of input keyword(s)/text is presented in Figure 1. *MSA* utilizes *Wikipedia* as a conceptual knowledge source. First, we build a search index of all *Wikipedia* articles. Second, we utilize association rule mining to uncover implicit associations between concepts by mining the *"See Also"* link graph of indexed articles. Third, we use the inferred association rules to construct a knowledge graph of concepts and their associations. Fourth, given a seed text, we map that text to the concept space by first generating a candidate set of concepts from the search index (explicit concepts), and then augmenting that set with strongly associated concepts from the knowledge graph (implicit concepts). Steps 1-3 are done offline making conceptual mapping very efficient at runtime.

Applications

We have a Web application of our framework to demonstrate typical use cases which map to real-world cognitive tasks that practitioners deal with today. These include a user initiated and interactive related-technology exploration tool, a prior art search application, and a company patent portfolio summarization functionality.

Due to limited space, we demonstrate in detail only one application of our framework to assist practitioners in technology exploration and landscape analysis.

Consider *"Cognitive Analytics" (CA)* as an example technology. Figure 2 (top) shows the concept space of *CA* by retrieving the top 10 most semantically related concepts using *MSA* (node size reflects association strength). We can clearly notice that: a) those concepts are very associated with *CA* as well as with one another, and b) they cover a wide spectrum of technological/conceptual landscape.

All patent data are indexed into our framework, consequently, we could facilitate landscape analysis of *CA*-related technologies by showing patenting and innovation progression as the ThemeRiver shown in Figure 2 (bottom). Each stream represents a patent class where stream width is proportional to the number of granted patents per class over 20

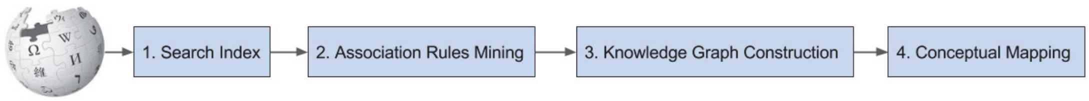

Figure 1: Concept space construction pipeline.

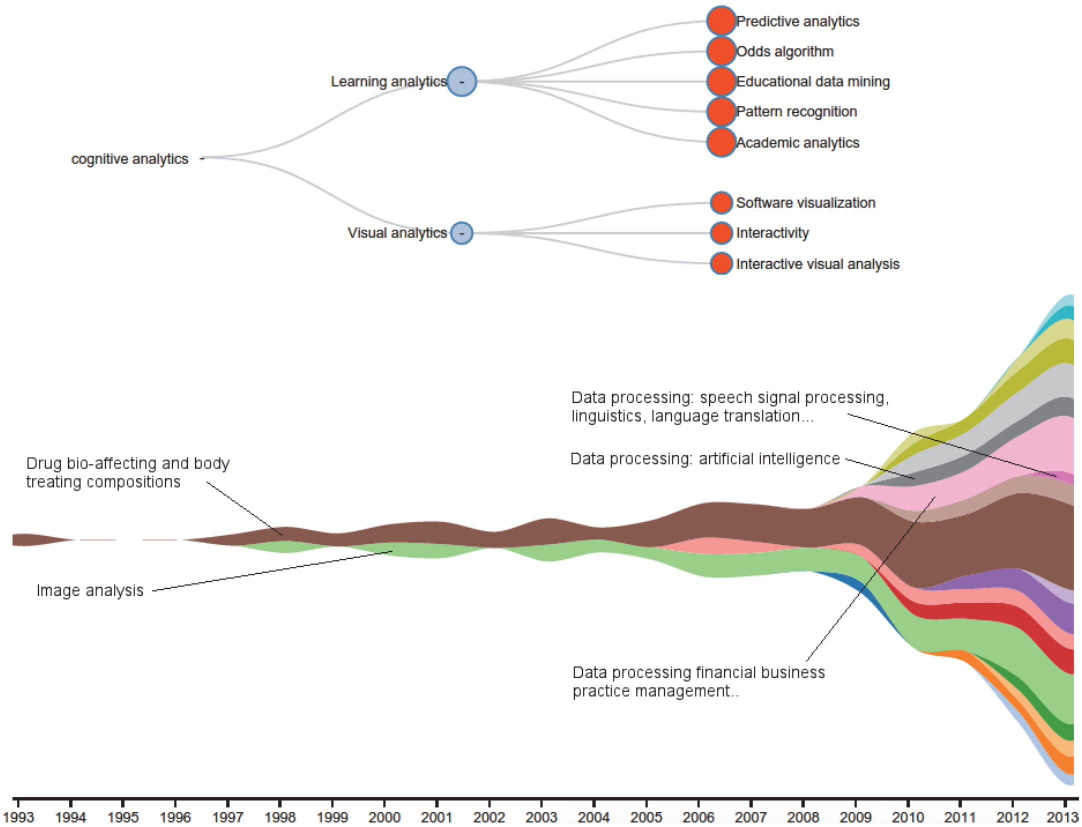

Figure 2: Top: Concept graph of *Cognitive Analytics*; explicit concepts are light blue nodes, and implicit concepts are red nodes. Bottom: ThemeRiver plot showing patenting evolution of *Cognitive Analytics* and related technologies in its concept graph.

years between 1993 and 2013. Streams allow practitioners to capture patenting trends that would otherwise require investigating huge number of patent documents in an iterative and time consuming manner. Those trends include:

- Patents were limited to two classes until 2005; ”*(382) image analysis*” (light green), and *”(514) drug bio-affecting and body treating compositions”* (dark brown).
- By 2010, granted patents expanded to cover various patent classes; e.g., ”*(706) data processing: artificial intelligence*” (dark gray) and ”*(705) data processing financial business practice management...*” (pink).
- A new application domain of *CA*-related technologies emerged in 2013; ”*(706) data processing: speech signal processing, linguistics, language translation...*” (purple).

Our framework encourages human-in-the-loop processing by providing users with interactive visualizations that support their cognitive tasks. For example, users can easily interact with the visualizations provided in Figure 2 by controlling the conceptual association strength, pruning possibly irrelevant concepts, and zooming in each stream to navigate through individual patents under that stream.

Another application area of our framework is patentability and prior art search. Through *MSA*'s conceptual mappings, we offer the practitioner an interactive and visual method of concept based navigation through relevant literature without manual query expansion or keyword based search. This is likely to offer a significant boost to the cognitive efficiency of the search and review process, translating into improvement in the quality of granted patents. Initial results suggest a relatively high recall rate of relevant prior art using *MSA*.

Summary

We present a semantic, visual, and interactive framework for innovation analytics powered by *MSA*. The framework supports many cognition and knowledge intensive tasks by offering a closed-loop iterative analysis of patent data and thus maximizes the potential for user exploration by providing a number of ”smart” starting-point semantically salient concepts making keyword search and exploration obsolete.