New Perspectives on Trademark Categorization using Text Analytics

Alex Lehmann (alexander.lehmann@ised-isde.gc.ca), Canadian Intellectual Property Office

As a national intellectual property office, the Canadian Intellectual Property Office (CIPO) is tasked with supporting the Canadian government's policies with respect to innovation and economic development. One government priority is to make Canada a world leader in clean technologies by enabling and supporting Canadian innovation in sectors such as renewable energy and low-carbon transportation. For CIPO, this means tailoring outreach and intellectual property (IP) services to firms involved in these sectors to support their growth and protect their innovations.

These tailored services include the registration of trademarks. As the vanguards of new and expanding products and enterprises, trademarks form a key component of any approach to executing the government's policies with respect to innovation. An important part of developing trademark services to support these policies is developing profiles of trademark users in the specific sectors of interest. The ubiquitous Nice Classification falls short in this respect as its categories are broader than the sectors in which the government wishes to develop firms profiles. For example, Class 4 of the Nice Classification contains trademarks to do with production of electrical energy; however, this class also contains products irrelevant to energy such as industrial waxes and leather oils. It is clear that in order to develop the firm profiles required for CIPO's objectives, a new approach to categorizing trademarks is warranted.

In this presentation, we introduce CIPO's ongoing research in employing analytical methodologies from text mining, machine learning, statistics, and physics literature to develop new trademark categorization schemes in support of Canadian government policy. We define the requirements for new trademark categorization systems and describe the data material from which they will be derived. We also discuss the use of data-driven methodologies to create these classifications as opposed to more traditional approaches.

We then introduce two promising aspects of our research to accomplish this task. In the first, we apply topic modeling, statistical learning, and complex network algorithms to cluster trademark filings into meaningful groups corresponding to industry sectors. This approach first uses a topic modeling approach based on non-negative matrix factorization to identify the subject matter components (lines of business, in the trademark context) making up each trademark application's text. We then embed this information in a complex network model and use community detection algorithms to identify associated lines of business; these groups of lines of business form the sectors we seek.

In the second approach, we use a tandem of neural network models to associate trademark filings with sectors as prescribed by the North American Industry Classification System (NAICS). Trademark applications are first analyzed by an deep-learning based natural language processing model, whose

output is then augmented with metadata from the application and firm characteristics. This intermediate data is then fed through another deep learning model that assigns a most likely NAICS sector to each filing.

We present preliminary results for each of the approaches discussed that are indicative of their potential to contribute to analytical efforts into trademark data. We also discuss next steps in the development of these methodologies and other potential use cases.

In conclusion, we find that data-driven trademark categorization schemes are useful tools to support the Nice Classification for trademark research. As technological advances continue in the field of data science, additional avenues to derive actionable intelligence from trademark data are paved and this data's value for informing government and business decision-making continues to climb. CIPO remains committed to advancing IP research using these data-driven tools to further Canadian government priorities.