Modeling Patent Clarity

Jonathan Ashtor

This article employs sophisticated machine learning techniques to model patent clarity. Using these models, unprecedented empirical analyses of the relationships between patent clarity and the public and private value of patent rights are conducted.

Specifically, I train a variety of machine learning models on a large dataset of U.S. patent applications that were rejected for failure of the "written description," "enablement" or "definiteness" requirements under 35 U.S.C. § 112. These rules generally test the form and language, rather than the substance, of a patent's disclosure, providing a useful proxy for clarity. I draw from the text of over 2 million published patent applications to construct an array of features regarding the disclosures and claims of these applications. The machine learning models trained on these features achieve 70% accuracy in predicting § 112 rejections.

Using these models, I first study how patent clarity relates to cumulative innovation. Patents serve a public function by disclosing new and useful technical information that other inventors can learn from. I find a significant positive correlation between patent clarity and cumulative innovation by third party inventors.

Next, I study the relationship between clarity and patent enforcement. Litigation theory teaches that unclear legal rights can lead to and exacerbate disputes. I observe both relationships empirically, finding that as patent clarity decreases, there is a significant increase both in the likelihood of enforcement actions and in the duration of these cases. These results empirically demonstrate the importance of patent clarity to the social and private functions of patents. They also serve to validate the methodology employed herein and pave the way for future in-depth empirical studies of patent clarity.