

Applying Natural Language Processing and Text Mining to Classifying Child Custody Cases and Predicting Outcomes

Sieh-Chuen Huang & Hsuan-Lei Shao *

Abstract

Recently there have been many studies of artificial intelligence that enable computers (machines) to simulate human thinking processes, assist human decision-making, and apply them to the field of law. However, most of previous studies have focused on analyzing the judges' thinking process and predicting the case outcomes, so that lawyers and the parties can refer to them. In order to obtain highly accurate results on analyzing cases or statutes, these studies must rely on legal experts to extract or retrieve key legal factors and code them manually. Based on the human coded data, the machine will build models and predict outcomes. On the other hand, this study attempts to adopt different methods. Instead of using manually coded data, we directly input legal texts which are in the form of unstructured natural language data (i.e., the original texts of court cases) into the machine, and observe whether the machine can successfully “understand” the judges' semantics and classify the cases. We collected 448 cases regarding child custody from 2012 through 2014. These parents were both Taiwanese and willing to acquire the custody, where the Taiwanese district court granted one parent sole custody. The machine used word segmentation techniques to build the Document Term Matrix. Next, we built the artificial neural network (ANN) model to classify the cases into two groups: father-sole-custody and mother-sole-custody. The model has a 77.25% overall accuracy and 0.8674 average F1 score on the

* Associate Professor, College of Law, National Taiwan University.

E-mail: schhuang@ntu.edu.tw

Associate Professor, Department of East Asian Studies, National Taiwan Normal University.

E-mail: hlshao@ntnu.edu.tw

testing data set. This confirms that the machine can "read" the legal texts to some extent and classify it. Since the speed of the machine is much faster than that of humans, this result, if being used in the legal data search system, will allow people to find the information (for example, to find the cases where the mother receives sole custody) more efficiently without the bother to rely on manual searching, reading, and selection of the most relevant cases. This research will also contribute to "human-machine collaboration" to support human decision-making, which is exactly the goal of legal data analytics in recent years.

Keywords: legal analytics, custody case, artificial neural network, text mining, human-machine collaboration