Algorithm for classification of employment hierarchy from electronic communications

Technology #cu15188

During most litigation procedures, the electronic discovery (eDiscovery) process is required so that opposing attorneys can share pertinent, non-confidential documents related to the case at hand. Confidential documents excluded during this phase often concern client-attorney communication and other correspondence related to legal advice. Traditionally, the document discovery process has been performed by legal professionals who must manually inspect thousands of documents, a process that is both tedious and expensive, especially since attorneys are typically billed by the hour. Unfortunately, professional titles can be fairly ambiguous and some organizations do not keep a comprehensive list of employees, support staff, and outside consultants. This technology is an algorithm that is able to classify the organizational role of a potentially unlisted employee based on previous email communications.

Algorithm automatically identifies organizational role of an employee using email history

With the average lawsuit involving approximately 200GB of data per party, there are potentially millions of different individuals involved in the sending or receiving of emails for each case. Often, it may be difficult for a particular organization to provide a comprehensive list of employees and job titles, especially for previous years. Furthermore, many organizations interact with external legal support and consultants which may not be efficiently catalogued by the organization. This technology employs a semi-supervised machine learning technique to construct a hierarchal network for each organization based on email communication. Using this method, the organization role of an unlisted employee may be inferred based on similar interactions with known employees. This technology will help improve existing methods of automated eDiscovery by performing a faster and more targeted search.

Initial evaluation of the algorithm was performed versus a gold standard eDiscovery algorithm. Using only 10% of the total labeled data, this technology was able to predict the organizational roles of unlabeled employees better than gold standard author-topic (AT) models.

Lead Inventor:

Apoorv Agarwal, Ph.D.
Applications:

• eDiscovery algorithm for law firms and legal service providers
• Internal IT system for employee management
• Public relations and business development outreach

Advantages:

• Incorporates network structure to identify hierarchal relationships
• Semi-supervised algorithm
• Requires only a small set of training data
• Superior performance compared to gold standard algorithm

Patent Information:

Patent Pending

Tech Ventures Reference: IR CU15188

Related Publications:


Inventors

Apoorv Agarwal