

WGBH - Reliability of Informant Cases

Client Contacts: Investigations Editor Paul Singer, Senior Investigative Reporter Philip Martin

Project Members: Chen Xin, Xiang Liu, Qitong Wang, Changhao Liang

Motivation & Background of the project

The WGBH Informant Project is an effort to create a comprehensive analysis of criminal cases involving organized crime which have been overturned due to the use of an informant in Massachusetts, New Hampshire, and Rhode Island over the past decade (2008-2019)

Goals, hypothesis, outcomes

Goals:

1. Extract criminal case data by scraping, using key search terms, all criminal case decisions where an informant was used from the appellate court websites.
2. Collect both full and partial reversed decisions in courts.
3. Uncover never before-seen patterns of judicial behavior and decision-making in criminal cases.
4. Recast and spur more aggressive news coverage of the court system and the use of informants.
5. Create a valuable understanding of the effectiveness of informant testimony.

Hypothesis:

In a criminal case, only the defendant has a right to an appeal in most states. Lower court cases can be challenged for any erroneous rulings by the judge on the evidence or on the law; the prosecutor's conduct during the case; or even the trial attorney's handling of your defense. Cases can be reversed in full or reversed in parts.

The client's data collection, analysis and newsgathering could reveal patterns based on the types of cases being overturned, the type of informant, the specific charge, whether it is a jury or non-jury trial, witness identification issues, errors in jury instruction, city or town where crime occurred or trial held, expert witness issues and much more.

Outcome:

Reliable Question Answering System. Creating a deep repository of searchable, online public information about the cases being overturned in Massachusetts and a credible quantification of those cases.

Non-goals, out of scope topics

1. Building a recommendation system based on our data model that could rank the reliability of informants
2. Developing graphical user interface for the final question and answering system

End result/product -> definition of done

1. We need to have converted all the text data, including court records and corresponding keywords, to usable quantitative data which can be used by some machine learning algorithms.
2. Do ablation studies when analysing the reliability of informants using more than 2 machine learning algorithms, analyzing which algorithm is more powerful and efficient to our problem.
3. Completion of the client-requested question-answering system and all project goals

Open questions, uncertainties

1. The quality of source data, which is crucial to the final results of our project.
2. The selection of key works, which may impact the effectiveness of our models
3. The reliability of our computational resources, since GPU could be required when we need intensive computation such as implementing a deep neural network.

Reference

We will scrape the following websites provided by clients to collect data:

<https://www.mass.gov/appellate-opinion-portal>

<https://www.courts.state.nh.us/supreme/opinions/>

<https://www.courts.ri.gov/Courts/SupremeCourt/Pages/Opinions%20and%20Orders%20Issued%20in%20Supreme%20Court%20Cases.aspx>

Our client will provide a list of key terms to use in the searching and scraping.