

# 智能法律系统项目

基于自然语言处理和因果统计推断实现法条推荐及上诉成功概率预测

---

发起人：朱晨阳，陈文奇

October 6, 2018

上海财经大学

# Table of contents

1. 智能法务系统背景及介绍
2. 本项目的立项依据
3. 初步实验分析
4. 项目计划及目标
5. 招贤纳才

## 智能法务系统背景及介绍

---

目前国外知名律所都开始使用智能法务系统，目前比较通用的智能法务系统主要聚焦于两大方面：

1. 法条的推荐系统。基于自然语言处理，通过输入案件信息，找到对应法条，方便律师查找法条，加快案件处理速度。
2. 为客户上诉进行预测。给予深度学习/机器学习，对案件上诉成功概率进行预测。

对于中国的法律智能系统主要有以下几个困难:

1. 中文的自然语言处理技术相对较弱, 对于中文语言中的分词、逻辑判断等方面还十分困难。
2. 我国法律体系相对独特, 法条构建还不够完善, 有大量解释文件、法条补充文件需要进行处理。因此需要大量的训练及样本。
3. 另外, 由于全国各地法官的判断标准还不够统一, 所以导致难以以统一标准进行预测上诉概率

## 本项目的立项依据

---

# 法律领域的智能系统符合目前宏观经济形势

首先，进一步开放政策，让更多的外国律所所有机会在北上广成立分公司，这些企业往往已具备了英语的智能法务系统，而没有中文的对应系统，使其竞争力在我国有一定限制。

其次，随着公民法律意识的逐渐增强，我们可以想象在未来的时间里，法律和维权意识会有显著提升。伴随而来的是整个律师乙方服务行业的进一步兴起。

最后，这一系统尚未被研发。

本项目针对目前法务系统的未满足需求和目前宏观经济形势，拟对以下议题进行分析：

1. 仅使用现有的自然语言处理方式，不在自然语言处理上进行大量研究。事实证明，此领域的研究已经超出了学生的范围。
2. 我们可以解决数据集的问题，通过法律文献查找网站以及爬虫，我们可以得到大量样本，并且能够对样本进行预处理，并形成推荐系统。
3. 我们可以解决上诉概率预测问题，事实证明，在我们的初步实验中这一概率的准确度已于65%左右。



## 初步实验分析

---

# 大量数据集保证充足样本进行训练

我们从法律网站上找到了2000条过去关于经济犯罪的数据，总数据库近百万条全国各地各类型案件。在初步实验中我们针对经济法犯罪问题进行了处理。该案件表头为共36列，包含以下信息：

标题、案号、案件类型、庭审程序、案由、文书类型、法院、判决日期、原告、被告、第三人法官、审判长、审判员、书记员、头部头部2、当事人、当事人2、庭审程序说明、庭审程序说明2、庭审过程、庭审过程2、庭审过程3、庭审过程4、庭审过程5、庭审过程6、法院意见、法院意见2、判决结果、判决结果2、庭后告知、庭后告知2、结尾、结尾2、附录

## 只选取五个变量+决策树模型即可得到63%乐观结果充分表明项目可行性

我们选取了以下五个变量:是否为终审,原被告分类,第三方分类,庭审程序分类,文书; 而Y是示性函数: 是否成功

```
from sklearn import tree
clf = tree.DecisionTreeClassifier()
clf = clf.fit(X[:-200],Y[:-200])

>>>(200 - np.sum(abs(clf.predict(X[-200:]) \
                    - Y[-200:]))) / 200

0.63
```

我们尚未使用自然语言处理或逻辑推断就已经取得0.63的成果。

## 项目计划及目标

---

## 以应用为主，不进行深度理论研究

本项目将作为学校挑战杯项目，不进行深度的理论研究，以应用机器学习和自然语言处理理论为主。我们最终期望能够达到预测结果至少在85%以上。

项目将分为以下几个阶段：

1. 数据集获取及预处理== 大约2个月。我们将选定两到三个法律领域，例如经济法、刑法等进行有针对性地处理。
2. 大量机器学习模型使用+大量运算== 大约3个月。使用大量决策树、SVM、逻辑回归、causal inference理论进行预测和分析。
3. 总结与论文== 1个月。应用+发表论文。

项目将在大四前完成。

以金融方面的数据集为基础，重点放在预处理中的逻辑推断、文本处理，以及之后的模型选取当中。方向为偏应用方面。旨在建立：

1. 智能法务推荐查找系统，帮助律师查找相对应法条。
2. 上诉成功概率预测。

招贤纳士

---

我们需要热爱机器学习、愿意付出的所有人！

需要大量的程序员和有idea有想法的所有人！！！！