Law Text Classification System Using AI with Recommendation Approach

Vinit Vijay Magre

Research Scholar, Department of Information Technology

MGM's Jawaharlal Nehru Engineering College, Aurangabad

Abstract- Our Mission is city without Crime and increasing trust about the police an online comprehensive crime reporting system to engage public, NGOs, police and government agencies to be greater brief, proactive and responsive to combat with crime and criminals. One of the aims of Artificial intelligence (AI) is the realization of natural dialogue between human beings and machines. In recent years, the dialogue structures, also called interactive conversational systems are the fastest developing area in AI. Many companies have used the dialogue systems technology to establish various kinds of Virtual Personal Assistants based on their applications and areas, such as Microsoft's Cortana, Apple's Siri, Amazon Alexa, Google Assistant, and Facebook's M. However, in this proposal, we have used the multi-modal dialogue systems which process two or more combined user input modes, such as speech, image recognition. Smart Virtual Assistant plays the important role to launching the FIR with speech to text conversion and after analyze the complaint apply appropriate laws with authenticate by Unique Identification (UID) for serious offense. As well as we are providing online web application to register non serious offense complaint. To enhance the communication between police and public which enables to improve the time usage for solving crimes and not an awful lot time is wasted to speak with police.

Keywords- Law Text Classification, Semi-supervised Learning, World Association Mining, Sentimental Analysis, Neural Network

I. INTRODUCTION

Now-a-days, many more users are willing to carry out online legal advice, legal case handling and other services in the media, which try to make several kinds of legal service patterns. How to find useful information quickly and accurately from legal cases description becomes very important for text classification. Text type

can help users to successfully deal with and take advantage of useful facts hidden in huge-scale documents. Hence, an intelligent online legal consulting platform is necessary that can automatically recognize and assign predefined criminal labels. In, a unified framework is proposed to expand short texts based on word embedding clustering and CNNs. In this paper we argue that law professionals would greatly benefit from the type of automation provided by deep learning. This is particularly the case of legal research, more specifically the preparation a legal practitioner has to undertake before initiating or defending a case.

The use of semi-supervised CNNs which directly learns text region embedding of unlabeled data, and then puts it into a supervised CNNs. At first, it needs to reduce the size of vocabulary. Because it is not necessary to use all the train text to gain good text classification performance, and that keeping lots of input text requires much resource. In the next place, it puts preprocessing text input into a simple model that merely includes convolution layers and top layer. The output of the simple model regards as part of CNNs input.

1.1 Motivation

- Motivated by the principle of compositionality, large multilayer neural network models have been employed for this task in an attempt to effectively utilize the law classifications.
- The semi-supervised CNN framework for text categorization that learns embedding's of text regions with unlabeled data and then labeled data.
- The semi-supervised framework learns a region embedding from unlabeled data and uses it to produce additional input (additional to one-hot vectors) to supervised CNN, where a region embedding is trained with labeled data.

II. REVIEW OF LITERATURE

Law Text Classification Using Semi-supervised Convolutional Neural Networks [1], This paper Conclude That, CNN architecture has been evaluated on a freely available large-scale data sets: the Chinese legal case description. We can show that semi supervised CNNs with tv-embeddings for text categorization improves performance compared with the traditional neural networks. Due to the limited space, this paper only considered the law text classification; therefore we will extend the system so that it is able to other applications, such as, traffic rules, film review, etc. Furthermore, the idea of the design proposed in this paper could also be enlightening.

Predicting judicial decisions of the European Court of Human Rights: a Natural Language Processing perspective [2], This paper presents the primary systematic take a look at on predicting the outcome of instances attempted by means of the European Court of Human Rights based entirely on text. We formulate a binary type assignment wherein the input of our classifiers is the textual content extracted from a case and the goal output is the actual judgment as to whether there has been a contravention of an editorial of the convention of human rights. Textual data is represented the use of contiguous phrase sequences, i.e., N-grams, and topics. Our fashions can are expecting the courtroom's choices with a strong accuracy 79%.

Next-Generation of Virtual Personal Assistants (Microsoft Cortana, Apple Siri, Amazon Alexa and Google Home) [3], This paper conclude that, Introduces the structure of Next-Generation of Virtual Personal Assistants that is a new VPAs system designed to converse with a human, with a coherent structure. This VPAs system has used speech, video and other modes for communication in both the input and output channel. Also, the VPAs system will be used to increase the interaction between users and the computers by using some technologies such as image/video recognition, speech recognition, and the Knowledge Base.

Sentimental Analysis [4], This paper concludes by, we have proposed a system that is used analyze the sentiment of the movie reviews using NLTK library and its functions present in python. The corpus can be taken from any trustable sources which can be static like we have used or dynamic like tweeter tweet analysis.

Crime area Detection and Crime area Record [5], In this paper we have overcome the problem of communication gap between the police during their investigation. The purpose of this paper is to develop an android application for crime area detection and store criminal records. It presents an application for the user that could offer an alternate direction for the users passing via crime area.

A Real-Time Crime Records Management System For National Security Agencies [6], This paper concluded that, The need for a computerized platform for a crime record management cannot overemphasized. The use of automated crime record management systems (CRMS) international to hold report of crime and criminals involved. Crime being an act against the law of a society is a threat to the well-being of the population and so, requires efficient and effective tracking.

Grid Based Authentication for Online Crime Reporting system [7], This paper conclude that, Online crime reporting System can be used to engage public and police works to be more quick, pre-emptive and reactive to flight with crime and criminals. In this paper we have proposed 2 systems 1. OCRS and 2. Grid based totally Authentication System to shield authentications from attacks like shoulder surfing and key loggers.

A Voice-Controlled Personal Assistant Robot [8], This paper concludes that, Voice commands given at the user end, are converted to text form using an on-line server in using real time speech signal processing. Speech commands converted to text commands are then transmitted to the robotic assistant the Bluetooth network of an Android based smart phone.

Crime Analysis and Prediction Using Data Mining [9], In this paper we have tested the accuracy of classification and prediction based on different test sets. Classification is done based on the Bayes theorem which showed more than 90 percent accuracy. Using this algorithm we trained various news articles and build a version. For testing we are inputting some test data into the model which shows better outcomes. Proposed system takes factors or attributes of a place and Apriori algorithm offers the frequent patterns of that area. The pattern is utilized for building a version for decision tree.

Mobile Application for Crime Recognition System [10], This paper conclude that, The proposed system can reduce the communication gap between the local police and the people (Users) and we created the communication bridge between them by keeping all the data on the remote server which can be accessible at any time by the both user and the police. This paper main advantage is that everyone can help the police by posting a crime immediately not only to the police but also aware among themselves and the post is verified by the police.

III. OPEN ISSUES

In the existing crime management system maximum of the operations are executed manually like taking action in opposition to the crimes, sending complaints, viewing status reports etc. If absolutely everyone desires to complain against the crimes he must visit to the police station. If we are doing the system physically, so many minor errors will occur. Detection within the preceding entries made and data cross verification is another essential function. This process might take extra time.

Disadvantages are:

- The current system is time consuming and no longer very user friendly.
- The officer dealing with a selected case can't take decision by way of himself even when he's having the first-hand knowledge/facts approximately the case and he can anticipate obstructions from higher government/officers.
- Even an efficient officer cannot/might not able to cope with more than one case at a time.
- In most of the cases, the innocent are accused in the previous system.
- The existing system could offer simplest investigation and there's no Advocating, Counseling facilities and many others.

IV. PROPOSED SYSTEM

In this paper, we design a law text classification with the characteristics of reaching a certain accuracy level. Thus, proposes a new semi-supervised CNN framework for text categorization that learns embeddings of text regions with unlabeled data and then labeled data. The embedding is a function that obtains low-dimensional feature that owns the predictive structure. Moreover, to improve the classification accuracy, we put forward the model based on CNNs. It makes use of CNNs to capture feature of the text region, and it applies feature vectors to text classification. And convolution layer converts input text to feature vectors, which employs the language model of N-Grams. We have made experiments with the English data sets law case descriptions. Empirical results validate the semi-supervised CNNs design and demonstrate its benefits, which can directly learn embeddings of text regions with unlabeled data and then put it into supervised CNNs.

The semi-supervised framework combines the following two steps.

1. Tv-embedding learning: Train a neural network u to predict the context from every region of size p so that u's convolution layer generates feature vectors for every textual region of size p for utilize in the

classifier in the top layer. It is this convolution layer, which embodies the tv-embedding that we switch to the supervised learning method within the next step.

2. Final supervised learning: Integrate the learned tv-embedding (the convolution layer of u) into β, so that the tv-embedded regions (the output of u's convolution layer) are used as an additional input to β 's convolution layer. Train this final model with labeled data.
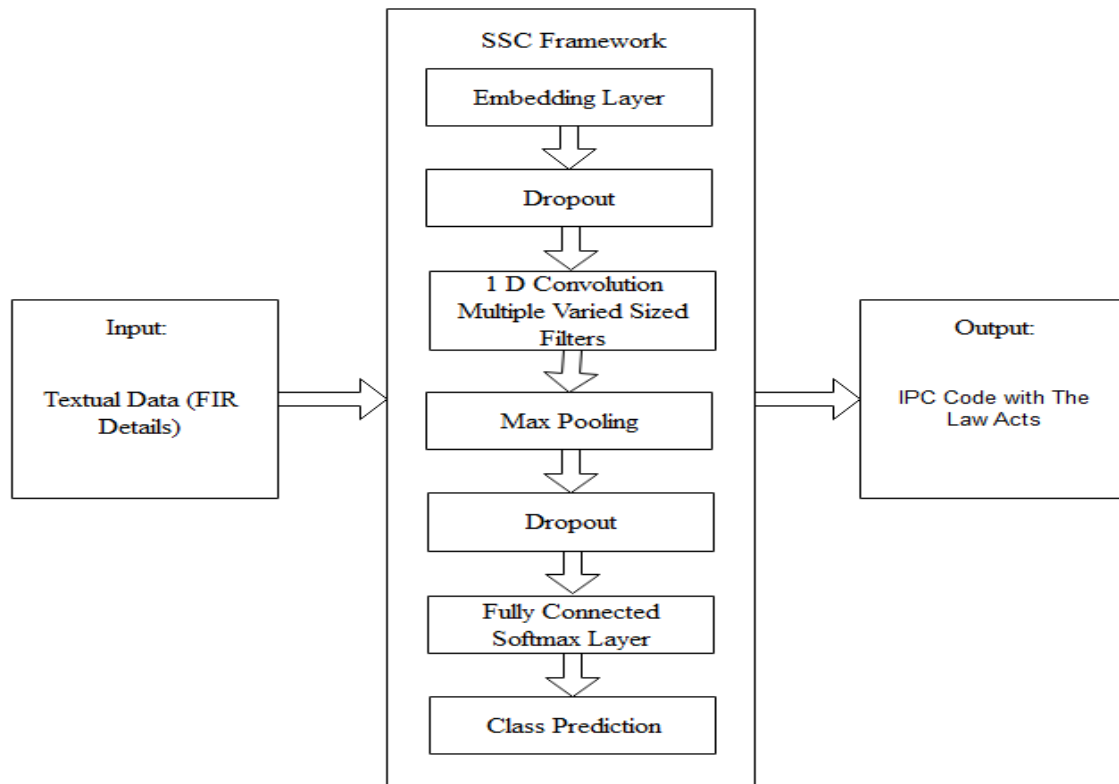
A. *Architecture*

Fig. 1 Proposed System Architecture (Source: Author)

Advantages are

- The semi-supervised CNNs with tv-embeddings for text categorization improve performance compared with the traditional neural networks.

- A semi-supervised CNN framework for multilabel text classification, which effectively learns the semantic information of test data.

- It helps to analyze the complaint and apply appropriate law to making the strong charge-sheet and taking quick action by police.

- It offers a common platform between police and general public to share crime related information.

B. *Mathematical Model*

1-D Convolutions over text:

Given a sequence of words $\omega_{1:n} = \omega_1, \ldots, \omega_n$, where each is associated with an embedding vector of dimension d. A 1D convolution of width-k is the result of moving a sliding-window of size k over the sentence, and applying the same convolution filter or kernel to each window in the sequence, i.e., a dot-product between the

concatenation of the embedding vectors in a given window and a weight vector u, which is then often followed by a non-linear activation function g.

Considering a window of words $\omega_i, \omega_{i+1}, \dots, \omega_{i+k}$ the concatenated vector of the ith window is then:

$$x_i = [\omega_i, \omega_{i+1}, \dots, \omega_{i+k}] \in R^{k \times d} \ \dots\dots (1)$$

The **convolution filter** is applied to each window, resulting in scalar values r_i, each for the ith window:

$$r_i = g(x_i \cdot u) \in R \ \dots\dots (2)$$

In practice one typically applies more filters, $u_1, \dots, u_l$, which can then be represented as a vector multiplied by a matrix U and with an addition of a bias term b:

$$r_i = g(x_i \cdot U + b) \ \dots.. (3)$$

with $r_i \in R^l$, $x_i \in R^{k \times d}$, $U \in R^{k \times d \times l}$ and $b \in R^l$

V. RESULT AND DISCUSSIONS

Experiments are done by a personal computer with a configuration: Intel (R) Core (TM) i3-2120 CPU @ 3.30GHz, 4GB memory, Windows 7, MySQL 5.1 backend database and jdk 1.8. The application is web application used tool for design code in Eclipse and execute on Tomcat server. The real time FIRs collection from every Police Station dataset of this application using applicant details.

The convolutional neural network is used for FIR content classification with the help of law IPC code and sections. The CNN algorithm performance is displayed in fig. 2.
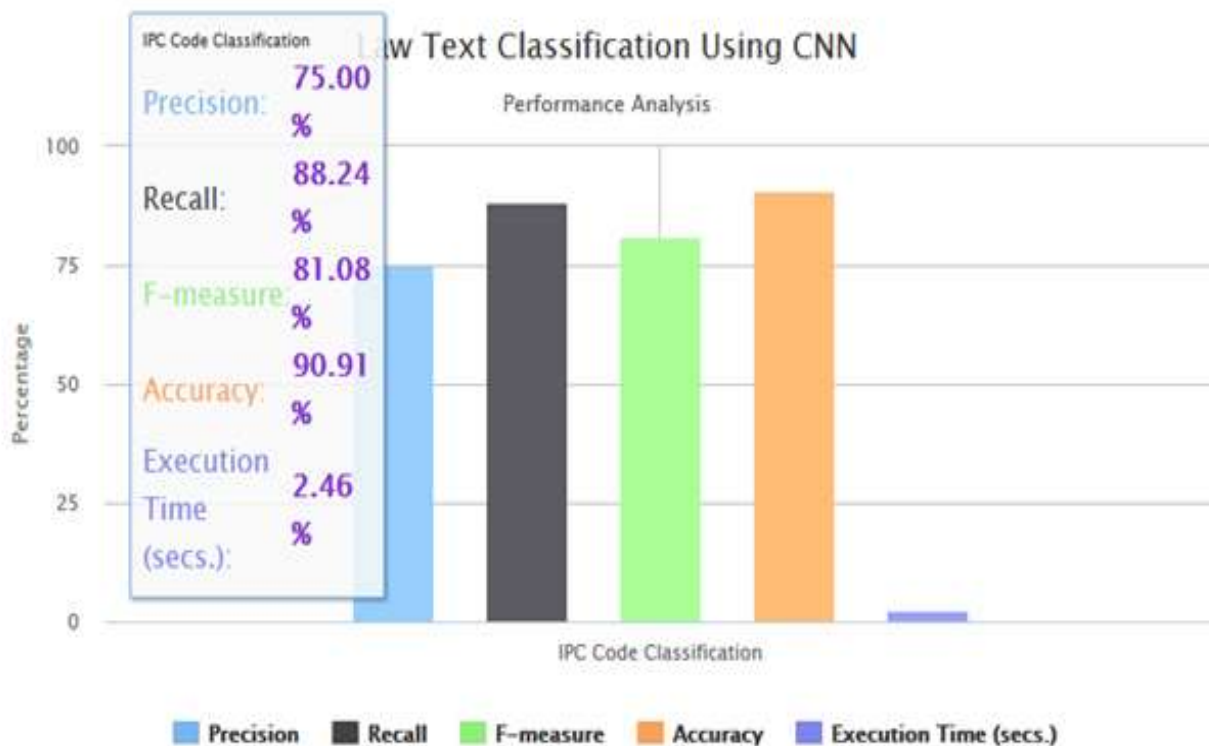

Fig. 2 Performance analysis of FIR law text classification (Source: Author)

The fig. 2 shows the accuracy is 90.91% law text classification as well as the execution time is 2.46 seconds.

VI. CONCLUSION

This paper proposed a new semi-supervised CNN framework for text categorization that learns embeddings of text regions with unlabeled data and then labeled data. We generate region vectors of the convolution layer, that is, conversion of bag-of-n-gram vectors to region vectors is done by a convolution layer. The semi-supervised CNN (SSC) framework improves performance of law text classification which gives output as IPC codes with law acts. This project offers a common platform between police and general public to share crime associated statistics. This project is helpful for analyze the complaint and apply appropriate law to making the strong charge-sheet and taking quick action by police. Future work is to implement the recursive convolutional neural network for sentiment analysis on text classification.

References:

[1] P. Li, F. Zhao, Y. Li and Z. Zhu, "Law text classification using semi-supervised convolutional neural networks," *2018 Chinese Control And Decision Conference (CCDC)*, Shenyang, 2018, pp. 309-313.

[2] Aletras, Nikolaos, Dimitrios Tsarapatsanis, Daniel Preotiuc-Pietro and Vasileios Lampos. "Predicting judicial decisions of the European Court of Human Rights: a Natural Language Processing perspective." PeerJ Computer Science 2 (2016): e93.

[3] V. Këpuska and G. Bohouta, "Next-generation of virtual personal assistants (Microsoft Cortana, Apple Siri, Amazon Alexa and Google Home)," *2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC)*, Las Vegas, NV, 2018, pp. 99-103.

[4] "Sentimental Analysis", 2015 International Journal of Emerging Technologies in Engineering Research (IJETER) Volume 6, Issue 4, April, 2018

[5] Aanchal Dabhere, Aniruddha Kulkarni, Ketaki Kumbharkar, Vrushali Chhajed, Sneha Tirth, "Crime Area Detection and Criminal Data Record". IJCSIT, Vol. 6 (1), 2015, 510-513

[6] Oludele Awodele, Onuiri Ernest E., Olaore Olufunmike A., Sowunmi Oluwawunmi O. Ugo-Ezeaba Anita A., "A Real-Time Crime Records Management System For National Security Agencies" European Journal of Computer Science and Information Technology Vol.3, No.2, pp.1-12, May 2015

[7] "Grid Based Authentication for Online Crime Reporting system", Naikwade Varsha et al, International Journal of Computer Science and Mobile Computing, Vol.4 Issue.3, March 2015

[8] "A Voice-Controlled Personal Assistant Robot", 2015 International Conference on Industrial Instrumentation and Control(ICIC) College of Engineering Pune, India, May 28-30,2015

[9] "Crime Analysis and Prediction Using Data Mining", First International Conference on Networks Soft Computing, IEEE, 2014

[10] "Mobile Application for Crime Recognition System", Volume V, Issue III, IJLTEMAS, ISSN 2278 2540, May 2016