www.ijaceeonline.com ISSN: 2456 - 3935
International Journal of Advances in Computer and Electronics Engineering
Volume 5, Issue 6, June 2020, pp. 1–9

A Survey on the Effective Machine Learning Approaches for Rhetorical Role Labeling of Sentences in Legal Documents

Divya Mohan
Research Scholar, Division of Computer Science and Engineering,
School of Engineering, Cochin University of Science and Technology, India
Email: div.mohan01@gmail.com

Dr. Latha Ravindran Nair
Associate Professor, Division of Computer Science and Engineering,
School of Engineering, Cochin University of Science and Technology, India
Email: latha5074@gmail.com

Abstract: *Rhetorical role labeling of sentences in a legal document refers to the process of understanding what semantic function a sentence is associated with, such as facts of the case, arguments of the parties, statute, precedent, the final judgment of the court, and so on. Rhetorical structure analysis has high-impact applications in natural language processing, for instances, text summarization, case law analysis, sentiment analysis, question answering, semantic search, etc. The output structures of the analysis contain high-level relationship between clauses and so provide valuable information and is highly beneficial for court document processing systems. Because of a wide range of applications and the necessity for automatic court document processing, automatic rhetorical structure analysis has been an area well noticed in the legal domain. In this paper, we propose to provide a survey of the various methods used for automatic rhetorical status classification of sentences in legal documents.*

Keyword: *court document processing, legal case documents, rhetorical status classification, rhetorical roles, semantic search, text summarization*

1. INTRODUCTION

Machine Learning refers to computer algorithms that can "learn" or improve in performance over time on some task. These algorithms have been successfully applied to automate various tasks that were once thought to necessitate human intelligence, for example language translation, fraud-detection, driving automobiles and facial recognition. If performing well, machine learning algorithms can produce automated results that approximate those that would have been made by humans. The success rate and superior performance have made these class of algorithms very popular.

Legal domain is yet another application area of machine learning technique which is currently gaining importance. One of the main reasons for this is the increasing accessibility of large legal corpora and databases. Still there is a noticeable gap in how much the state-of-the-art techniques are being incorporated in the legal domain, as legal practice is thought to require advanced cognitive abilities.

But there are various legal tasks such as predicting the outcomes of legal cases, finding hidden relationships in legal documents and data, electronic discovery, and the automated organization of documents which can be done using machine learning algorithms. By achieving this goal, a multi-disciplinary community can be built that can benefit from the competencies of both law and computer science experts.

Very little innovation in terms of technology is witnessed by Indian legal sector as the lawyers still feel comfortable and rely on the methods and methodologies that were designed years ago. But recent trends in artificial intelligence and especially machine learning have opened various opportunities in the field of Indian legal research. The volume and versatile nature of Indian legal system can aid legal practitioners to get unparalleled insight into the legal domain with the proper usage of machine learning algorithms. It can provide lawyers with highly efficient and advanced tools helping lawyers become better in advising clients or litigating.

Cite this paper:

Divya Mohan, Latha Ravindran Nair, "A Survey on the Effective Machine Learning Approaches for Rhetorical Role Labeling of Sentences in Legal Documents", International Journal of Advances in Computer and Electronics Engineering, Vol. 5, No. 6, pp. 1-9, June 2020.

Over a previous couple of years, high level of competition is witnessed globally within the legal industry.Thus, it becomes inevitable for the law industry to realize this competition and make use of the advancements in technology to meet client requirements. If these changes are not realized then it is sure that those law firms will be obsolete within the next few years.

Discourse analysis has high-impact applications in natural language processing, for instance, text summarization [17], [8], sentiment analysis [6], and question answering [9]. The output structures of the analysis contain high-level relationship between discourses and so provide valuable information to such tasks. In addition to the wide range of applications and the necessity for automatic court document processing, automatic rhetorical structure analysis has not been well noticed in the legal domain.

The most important part of a lawyer's or law student's reading matter are law judgments. In order to make judgments accessible and to enable rapid scrutiny of their relevance, they are usually summarised by legal experts. These summaries vary according to target audience (e.g. students, solicitors). Manual summarisation can be considered as a form of information selection using an unconstrained vocabulary with no artificial linguistic limitations. Automatic summarisation, on the other hand, has postponed the goal of text generation and currently focuses largely on the retrieval of relevant sections of the original text. The retrieved sections can then be used as the basis of summaries with the aid of suitable smoothing phrases.

Rhetorical labelling of sentences in legal judgments means what semantic function a sentence is associated with, such as facts of the case, arguments of the parties, statute, precedent, the final judgement of the court, and so on. It has wide application in the task of text summarization of legal documents. Rhetorical status classification also finds its application in areas such as court document processing systems, sentiment analysis, question answering, semantic search [14] etc. However, legal case documents are usually not well structured [15,18], and various themes often interleave with each other. For instance, the reason behind the judgment (Ratio of the decision) often interleaves with Precedents and Statutes. Hence it sometimes becomes difficult even for human experts to understand the intricate differences between the rhetorical roles. Hence, automating the identification of these rhetorical roles is a challenging task.

The re-emergence of deep learning, as the go-to technique for most AI-driven applications had a great impact in the area of natural langage processing. The clear reason for this is that deep learning has repeatedly demonstrated its superior performance on a wide variety of tasks including speech, natural language, vision, and playing games. On similar grounds, usage of deep learning approaches had played a major role in rhetorical status classification of legal documents.

This paper is organized into various sections. In section 2, review of various works carried out in this area are listed. Section 3 gives the findings of the review conducted. And Section 4 concludes paper and lists the future scope of this area of study.

2. LITERATURE SURVEY

Various works were carried out in this area, that employed machine learning methods. These methods used hand-crafted features, such as linguistic cue phrases indicative of a rhetorical role [14], the sequential arrangement of labels [15], and so on. Some of these features, e.g., indicator cue phrases, are largely dependent on legal-expert knowledge which is expensive to obtain. Also, the hand-crafted features developed are often specific to one or a few domains/categories (e.g., Cyber crime and Trade secrets [19]). It has not been explored whether one can device a set of features that works for documents across domains.

In this paper, we give a survey of the works which used various machine learning approaches for handling the task of labelling the sentences in legal case documents with rhetorical roles, in particular emphasizing the use of deep learning models for this task.

2.1 Role Identification using Conditional Random Fields

M.Saravanan. et al. [10] described a method for automatic identification of rhetorical roles in legal judgments based on rules and machine learning algorithms. They used manually annotated sample documents on three different legal sub-domains namely rent control, income tax and sales tax. They trained an undirected graphical model to segment the documents along different rhetorical roles. The features used for this work are namely cue words, state transition, named entity, position and other local and global features. The main purpose of segmenting texts with identified roles is in the re-ordering of sentences used for text summarization. The important sentences for summarization are extracted based on term distribution model. In this work, a fixed set of seven rhetorical categories based on Bhatia's (1993) genre is used, as listed in Table 1.

In this evaluation, human annotated documents were matched to test successfully the performance of the system, as it is a common norm in IR tasks to consider human performance as an upper bound.

The evaluation measure used in this work to compare the inter-agreement between sentences extracted by two human annotators for role identification in legal judgments is Kappa. The Kappa

ISSN: 2456 - 3935
International Journal of Advances in Computer and Electronics Engineering
Volume 5, Issue 6, June 2020, pp. 1–9

score of 0.803 shows the good reliability of human annotated corpus.

TABLE I RHETORICAL ANNOTATION SCHEME [10]

Roles	Description
Identifying the case	The sentences that are present in a judgment to identify the issues to be decided for a case. Courts call them as “Framing the issues”.
Establishing facts of the case	The facts that are relevant to the present proceedings/litigations that stand proved, disproved or unproved for proper applications of correct legal principle/law.
Arguing the case	Application of legal principle/law advocated by contending parties to a given set of proved facts.
History of the case	Chronology of events with factual details that led to the present case between parties named therein before the court on which the judgment is delivered.
Arguments (Analysis)	The court discussion on the law that is applicable to the set of proved facts by weighing the arguments of contending parties with reference to the statute and precedents that are available.
Ratio decidendi (Ratio of the decision)	Applying the correct law to a set of facts is the duty of any court. The reason given for application of any legal principle/law to decide a case is called Ratio decidendi in legal parlance. It can also be described as the central generic reference of text
Final decision (Disposal)	It is an ultimate decision or conclusion of the court following as a natural or logical outcome of ratio of the decision

Their work results shows that CRF-based and rule based methods perform well for each role categories compared to other methods till date. CRF-based method performs extremely well and paired t-test result indicates that it is significantly (p < .01) higher than other methods on rhetorical role identification for legal judgments belonging to rent control, income tax and sales tax sub-domains.

2.2 Theme Identification for Text Summarizer

A. Farzindar. et al. [4] used a methodology to identify the thematic structure to use the same for extracting important sentences from a document. The identification of themes separates the key ideas from the details of a judgment and improves readability and coherency in the summary. This work was done on a corpus of 3500 judgments of the Federal Court of Canada. Four themes were identified which divide the legal decisions into thematic segments as given in Table 2, based on the experimental work of judge MailHot [1].

TABLE II THEMES AND DESCRIPTIONS [4]

Theme	Description
Introduction	describes the situation before the court and answers these questions: who? did what? to whom?
Context	explains the facts in chronological order, or by description. It recomposes the story from the facts and events between the parties and findings of credibility on the disputed facts.
Juridical analysis	describes the comments of the judge and finding of facts, and the application of the law to the facts as found.
Conclusion	expresses the disposition which is the final part of a decision containing the information about what is decided by the court.

The information used for thematic segmentation are presence of significant section titles, positions of segment, identification of direct or narrative style, certain linguistic markers, etc. The evaluation of the summarizer shows that it has superior performance and obtained 90% correct segmentation for thematic segmentation module.

2.3 Facts and Principles Identification

O.Shulayeva. et al. [15] proposes a work that provides a novel, preliminary contribution towards automated identification of legal principles and facts embedded within common law citations. This work is motivated from the doctrine of stare decisis, which can be translated from Latin as to ‘stand by the decided cases’, where a case under consideration that has facts similar enough to precedent cases should receive similar decisions as the precedents. Citations from existing case law are used to illustrate legal principles and facts that define the conditions for application of legal principles in the current case. Citation analysis can help legal practitioners to identify which principles have applied in a certain case and which facts have been selected as the

ISSN: 2456 - 3935
International Journal of Advances in Computer and Electronics Engineering
Volume 5, Issue 6, June 2020, pp. 1–9

'material' facts of the case, i.e. the facts that influenced the decision and which are crucial in establishing the similarity between two cases.

To support citation analysis, existing electronic tools, such as electronic databases, provide one-word summaries for relationships between cases (e.g. 'applied'). However, it does not extract information about the facts and legal principles of cited cases. Thus, the readers are expected to understand the full text and identify the applicable law and the correct way to use it. This work can be considered as a source of aid to citation analysis. Their work aims to apply machine learning methodology in order to automatically identify legal principles and facts associated with case citations.

A gold standard corpus is created, with sentences containing cited legal principles and facts manually annotated. The corpus for the gold standard was compiled from 50 common law reports that had been taken from the British and Irish Legal Institute (BAILII) website in RTF format. For the annotation task, annotation guidelines – high level task definition, descriptions and examples for each category and analysis of cases were created by Annotator 1 and Annotator 2 was trained to use the same. Based on the written guidelines annotators were expected to identify sentences that contained legal facts and principles of the cited cases. Sentences which had no principles or facts were annotated as neutral.

The results of the inter-annotator agreement study show an agreement percentage of 83.7. The intra-annotator agreement study showed that Annotator 1 was extremely consistent with an agreement percentage of 97.3. The gold corpus of 50 reports created by Annotator 1 was used for training a machine classifier, which could be used for automated annotation.

A Naive Bayesian Multinomial Classifier is then applied to the corpus using a set of linguistic features to automatically identify these sentences. The features selected are Part of speech tags, Unigrams, Dependency pairs, Length of the sentence, Position in the text, Cit—a feature which indicates whether there is a citation instance in the sentence.

The machine learning experiments were conducted using Weka, a collection of machine learning algorithms for data mining tasks. Results were reported using ten-fold cross validation method. The experiment shows good result with classifier identifying 85% of instances correctly and achieving Kappa score of 0.72.

The main results are a demonstration that (a) the human annotation task is feasible, i.e. human annotators can achieve reasonable agreement on which sentences in legal judgments contain cited facts and principles and (b) it is feasible to automatically annotate sentences containing such legal facts and principles to a high standard. The reported studies lay the basis for further applications, including creation of metadata for search and retrieval purposes, compilation of automated case treatment tables containing summaries about legal principles and material facts of cases, and automated analysis of reasoning patterns and consistency applied in legal argumentation.

2.4 Role classifier for Text Summarization

TABLE III RHETORICAL ANNOTATION SCHEME FOR LEGAL JUDGMENTS [3]

Label	Description
FACT	The sentence recounts the events or circumstances which gave rise to legal proceedings. e.g. *On analysis the package was found to contain 152 milligrams of heroin at 100% purity.*
PROCEEDINGS	The sentence describes legal proceedings taken in the lower courts. e.g. *After hearing much evidence, Her Honour Judge Sander, sitting at Plymouth County Court, made findings of fact on 1 November 2000.*
BACKGROUND	The sentence is a direct quotation or citation of source of law material. e.g. *Article 5 provides in paragraph 1 that a group of producers may apply for registration.*
FRAMING	The sentence is part of the law lord's argumentation. e.g. *In my opinion, however, the present case cannot be brought within the principle applied by the majority in the Wells case.*
DISPOSAL	A sentence which either credits or discredits a claim or previous ruling. e.g. *I would allow the appeal and restore the order of the Divisional Court.*
TEXTUAL	A sentence which has to do with the structure of the document or with things unrelated to a case. e.g. *First, I should refer to the facts that have given rise to this litigation*
OTHER	A sentence which does not fit any of the above categories. e.g. *Here, as a matter of legal policy, the position seems to me straightforward*

ISSN: 2456 - 3935
International Journal of Advances in Computer and Electronics Engineering
Volume 5, Issue 6, June 2020, pp. 1–9

B.Hachey. et al. [3] proposes a classifier that determines the rhetorical status of sentences in texts from a corpus of judgments of the UK House of Lords. They have gathered a corpus of 188 judgments from the years 2001–2003 from the House of Lords website.

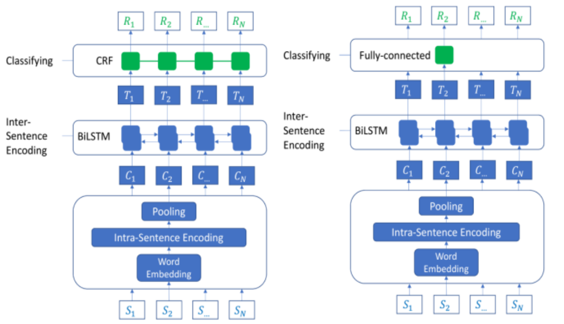
The annotation task was done by two annotators based on the guidelines developed by one of the authors, one of the annotators and a law professional. Inter-annotator agreement measure was calculated using Kappa coefficient of agreement with a score of 0.83. In this work, linguistic analysis is done to compute information to be used to provide features for sentence classifier. The various steps in linguistic analysis are Lemmatization, Named Entity recognition, Chunking and Clause Identification followed by verb and subject features.

The features used for experimenting in this work includes Location, Thematic Words, Sentence Length, Quotation, Entities and Cue Phrases. Four classifiers were used namely C4.5 decision trees, Naive Bayes (NB), Winnow algorithm and Support Vector Machine using polynomial kernels.

The experimental results reported in this paper were obtained using 10-fold cross validation over the 40 documents. In this work, the actual per-sentence (micro-averaged) F-score improvement is relatively high achieving an improvement of between 29.4 and 53.4 points.

2.5 Rhetorical status classification using Deep Learning Model in UK Legal Documents

Vu D. Tran. et al. [19] describes the approach of rhetorical role labeling of sentences in legal documents using deep learning models. Deep learning has shown effective in natural language processing tasks yielding promising results. Their work is done under the assumption that rhetorical label of a sentence not only depends on the content of the sentence but also on its relationship with other sentences. So inter-sentence dependency and intra-sentence dependency are considered during relationship modeling.



(a) Classifying layer using CRF (b) Classifying layer using Fully-Connected

Figure 1 Rhetorical Status Classification Models [19]

This work uses a deep learning model as shown in Figure 1. The various layers in the neural network model are Word Embedding, Intra-sentence Encoding, Pooling, Inter-sentence Encoding and Classification layer.

- Word embedding - maps a word into a continuous vector space. In this work, the authors had employed GloVe [11] for word embedding. This helps in capturing the contextual similarity among words – words which are frequently used in same contexts are near to each other in vector space.
- Intra-sentence Encoding – encoding the local information of a sentence is done either by using a BiLSTM or CNN. BiLSTM encodes the temporal information in a sentence both from left to right and right to left. CNN can be used for encoding by applying convolutional operations on word gram which captures the local information within a sentence.
- Pooling- extracts the required features by transforming the varying length sentences into fixed length vector. Max-over-time pooling is used as described by Kim et al. [12].
- Inter-sentence Encoding – the temporal dependency of the input sentence sequence is encoded using BiLSTM which helps to consider the information from the beginning and end of the sequence at decision time as it is bi-directional.
- Classification – two options (Figure 1(a) and (b)) were employed for classifying layer.

(i) In natural language processing tasks like sequential tagging [13], [16], CRF is often used on top of BiLSTM for predicting sequentially dependent labels. BiLSTM puts dependency to the features of each sentence by binding one sentence with others in temporal order, while CRF captures dependency between the rhetorical status outputs. Thus, using BiLSTM + CRF, we capture dependency both in terms of features and outputs.

(ii) Fully Connected (FC) Layer is used to predict labels independently. When FC layer is used on top of inter-sentence encoding layer, it binds one sentence with others in temporal order and thus prediction of output label of one sentence is influenced by the information from other sentences. When inter-sentence encoding layer is removed, it results in completely independent prediction.

The dataset used in this work is the one used in [3]. It is a collection of 40 judgments of the House of Lord from 2001 to 2003. The dataset consists of 10,169 sentences annotated by 2 annotators with agreement of 0.83 Kappa co-efficient. The experiments were carried out in different settings such as hidden size of 300, 500 and 1000 were used in each inter-sentence encoding and intra sentence encoding BiLSTM layers. The classification layer was experimented with both CRF and FC layers. Among all the experimental settings

[BiLSTM (1000) + BiLSTM (1000) + FC]

i.e. model with BiLSTM hidden size of 1000 for both intra and inter sentence encoding and FC for classifying layer yielded the best performance with F-score of 68.6%.

2.6 Rhetorical status classification using Deep Learning Model in Indian Legal Documents

P.Bhattacharya. et al. [20] uses deep learning models to identify rhetorical roles of sentences in Indian legal documents. This work illustrates in detail about annotation study carried out in labeling documents and details about the deep learning model used for automating this task. Two deep learning models are explored in this work to automate the task of rhetorical role labeling namely a Hierarchical BiLSTM model and a Hierarchical BiLSTM-CRF model.

Legal judgments from the Supreme Court of India is used as dataset in this work. They used deep learning models for supervised classification across seven rhetorical labels (classes) and over 50 documents from five different legal domains :

(i) Criminal – 16 documents
(ii) Land and property – 10 documents
(iii) Constitutional– 9 documents
(iv) Labour and Industrial – 8 documents
(v) Intellectual Property Rights – 7 documents.

The task of role identification is challenging for Indian Case docs mainly because they are poorly structured – rhetorical roles often interleave, and sentences of the same role may be at different positions in different documents. Another reason for the challenge is the absence of paragraph headings – the presence of which can aid as useful signals for rhetorical roles.

Three senior law students from the Rajiv Gandhi School of Intellectual Property Law, India were the annotators. Seven rhetorical roles as listed in Table -4 were identified in consultation with the annotators. The annotation process was done by annotators using GATE Teamware Tool [2]. Initially an annotation manual listing the guidelines to be followed for annotation was prepared. Each annotator was asked to annotate the document independently followed by a joint discussion to resolve any issues.

Inter Annotation Agreement measure was computed using pairwise Precision, Recall and F-Score which is better compared to Kappa score [7]. Suppose we have three annotators (A1,A2,A3), computing Inter Annotator Agreement (IAA) – average of pairwise agreements is taken – avg(A1&A2,A2&A3,A1&A3). Three types of agreement between two annotators are considered- Correct, Partial and Missing/Spurious. Correct type of agreement happens when both annotators marks the same span of text with the same label. Partial type of agreement is when two annotators marks slightly different span of text with the same label. Missing or spurious type of agreement is when two annotators marks same span of text with different labels.

TABLE IV RHETORICAL ROLES AND DESCRIPTION [20]

Role	Description
Facts (abbreviated as FAC)	This refers to the chronology of events that led to filing the case, and how the case evolved over time in the legal system (e.g., First Information Report at a police station, filing an appeal to the Magistrate, etc.)
Ruling by Lower Court (RLC)	Since we are considering Supreme Court case documents, there were some judgements given by the lower courts (Trial Court, High Court) based on which the present appeal was made (to the Supreme Court).The verdict of the lower Court and the ratio behind the judgement by the lower Court was annotated with this label.
Argument (ARG)	The Court's discussion on the law that is applicable to the set of proven facts by weighing the arguments of the contending parties.
Statute (STA)	Established laws, which can come from a mixture of sources –Acts, Sections, Articles, Rules, Order, Notices, Notifications, Quotations directly from the bare act, and so on.
Precedent (PRE)	Prior case documents. Instructions similar to statute citations.
Ratio of the decision (Ratio)	Application of the law along with reason- ing/rationale on the points argued in the case; Reason given for the application of any legal principle to the legal issue
Ruling by Present Court (RPC)	Ultimate decision / conclusion of the Court following from the natural / logical outcome of the rationale

ISSN: 2456 - 3935
International Journal of Advances in Computer and Electronics Engineering
Volume 5, Issue 6, June 2020, pp. 1–9

Precision = (correct+0.5*partial)/(correct + spurious + partial)

Recall=(correct+0.5*partial)/(correct+missing+partial)

F-Score = ((β^2+1) * Precision * Recall)) / ((β^2 * Precision) + Recall) [β=1]

GATE annotation tool computes three variants for each of the Precision, Recall and F-score measures – Strict measure – it considers all partial matches as in-correct or spurious. Lenient measure – considers all partial measures as correct and Average measure – which takes the average of the strict and lenient measures.

TABLE V AVERAGE INTER-ANNOTATOR AGREEMENT OF THE THREE ANNOTATORS IN TERMS OF F-SCORE A MEASURED BY GATE TOOL [20]

Labels	ARG	FAC	PRE	Ratio	RLC	RPC	STA
Strict	0.692	0.715	0.654	0.677	0.740	0.654	0.857
Lenient	0.953	0.934	0.878	0.908	0.925	0.968	0.967
Average	0.823	0.817	0.814	0.821	0.819	0.798	0.898

The standard data set was curated by the following process. In order to assign label to each sentence, they took a majority voting of the labels given by the 3 annotators. Pre-processing of documents was done prior to applying supervised machine learning models. Each of the 50 documents were split to form sentences using tool SpaCy (https://spacy.io/). There were 9,380 sentences and each sentence was considered as a unit for which one of the seven rhetorical labels will be assigned.

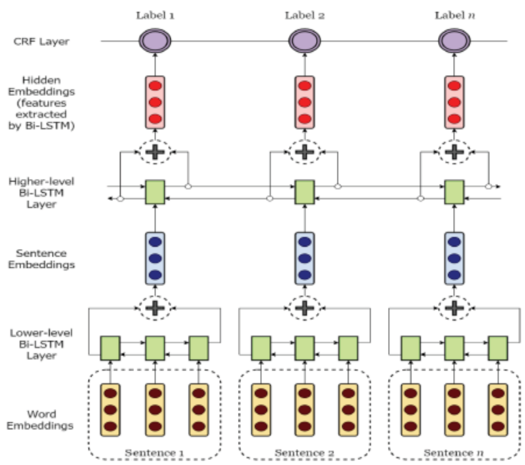


Figure 2 Proposed neural model -Hierarchical -BiLSTM-CRF [20]

They used hierarchical BiLSTM [5] for extracting features automatically to identify rhetorical roles. The sentence embeddings used to initialize the models were either randomly initialized word embeddings using another BiLSTM or pre-trained sentence embeddings from the set of documents of the same domain. Another add-on was also experimented with a CRF layer on top of the BiLSTM for taking into account label dependencies. They used hierarchical BiLSTM [5] for extracting features automatically to identify rhetorical roles.

The sentence embeddings used to initialize the models were either randomly initialized word embeddings using another BiLSTM or pre-trained sentence embeddings from the set of documents of the same domain. Another add-on was also experimented with a CRF layer on top of the BiLSTM for taking into account label dependencies. They used hierarchical BiLSTM [5] for extracting features automatically to identify rhetorical roles. The sentence embeddings used to initialize the models were either randomly initialized word embeddings using another BiLSTM or pre-trained sentence embeddings from the set of documents of the same domain. Another add-on was also experimented with a CRF layer on top of the BiLSTM for taking into account label dependencies.

Cross-validation, a standard way of evaluating machine learning models was used in this work. In particular they used 5-fold cross-validation on the 50 documents. In each fold, 40 documents were used for training and the rest 10 documents were used for testing the model. After 5 such folds, the result was averaged to give the performance measure. Macro averaged Precision, Recall and F-score were used for evaluating the performance of the algorithms. Hier-BiLSTM-CRF was inferred as the best performance model in this work which gave a F-score of 0.8208.

3. FINDINGS

This paper was written with the objective of conducting a survey on the different machine learning approaches used for rhetorical role labelling of sentences in legal documents. The survey mainly focussed on the better performance of works in which deep learning approaches were used.

The main reason for this is because deep learning models are nowadays highly used in natural language processing yielding better and superior results compared to the ones using classical machine learning methods. The task of rhetorical role labeling can be extensively used in less explored area of Indian legal domain. This will definitely be helpful to the public who can access the legal documents and have better understanding of the same, which would otherwise require legal help which is often expensive.

Table 6 lists the overview of the survey conducted by listing the title of work, algorithm and method used, dataset used and performance measure.

ISSN: 2456 - 3935
International Journal of Advances in Computer and Electronics Engineering
Volume 5, Issue 6, June 2020, pp. 1–9

TABLE VI COMPARISON OF THE WORKS ON RHETORICAL ROLE LABELING IN LEGAL DOCUMENTS

1	Title of Work: Algorithms and Methods Used: Dataset Used: Performance Measure:	- Automatic Identification of Rhetorical Roles using Conditional Random Fields for Legal Document Summarization - CRF method - 200 legal judgments up to year 2006 from the website www. keralalawyer.com - F-score:Rent Control Domain-0.849 Income Tax Domain - 0.817 Sales Tax Domain - 0.787-
2	Title of Work: Algorithms and Methods Used: Dataset Used: Performance Measure:	- LetSum, an automatic Legal Text Summarizing system - Table style summary formed by Thematic segmentation, filtering, selection and production stages - Legal record of the proceedings of federal courts in Canada - Thematic segmentation stage -90% coverage
3	Title of Work: Algorithms and Methods Used: Dataset Used: Performance Measure:	- Recognizing cited facts and principles in legal judgements - Naive Bayes Multinomial classifier - 50 common law reports from British and Irish Legal Institute (BAILII) website - F-score: Principles – 0.810, Facts -0.818
4	Title of Work: Algorithms and Methods Used: Dataset Used: Performance Measure:	- A Rhetorical Status Classifier for Legal Text Summarisation - C4.5 decision tree, Naive Bayes, Winnow algorithm, Support Vector Machine - 188 judgments from the years 2001–2003 from the House of Lords website - F-score: C4.5 (65.4- Location feature) NB (51.8-Quotations feature) Winnow (41.4 – Thematic words) SVM (60.6-Thematic words)
5	Title of Work: Algorithms and Methods Used: Dataset Used: Performance Measure:	- An Approach of Rhetorical Status Recognition for Judgments in Court Documents using Deep Learning Models - Neural network model with BiLSTM hidden size of 1000 for both intra and inter sentence encoding and FC for classifying layer - 40 judgments from the House of Lord website from 2001 to 2003. - F-score: 68.6%
6	Title of Work: Algorithms and Methods Used: Dataset Used: Performance Measure:	- Identification of Rhetorical Roles of Sentences in Indian Legal Judgments - Neural network model - hierarchical BiLSTM-CRF using PreTrained embeddings - 50 legal judgments from Supreme Court of India - F-score: 0.8208

4. CONCLUSION

Machine Learning algorithms are those that can "learn" or improve in performance over time on some task. These algorithms are efficiently used in various fields. The purpose of applying machine learning techniques in legal domain is to increase the overall performance of various tasks in legal systems. Rhetorical role labeling is the task of assigning labels to sentences based on the semantic function of a sentence in legal documents. Many machine learning techniques have been used for rhetorical role labeling. Such classical machine learning techniques often use hand-crafted features which are often specific to a domain. While deep learning models extract features on their own and their use in natural language processing yields promising results compared to classical machine learning approaches.

This paper is a survey conducted on the various works that used machine learning approaches for rhetorical role labeling. The survey covers works that uses classical approaches as well as deep learning neural models. A comparison of the various works is also given as part of the survey. Efficient Rhetorical role labeling can be effectively used for other legal tasks such as text summarization, sentiment analysis, question answering, etc. Automatic identification of rhetorical roles can make many legal tasks easier and effective, especially for the public to avail various use cases like argument recommenders, reasoning monitors, semantic viewers, semantic search, decision summarizers etc. [21].

The paper highlights the various works carried out in this area of legal domain. This work can be used as a reference to the existing works and inferences drawn to enhance the efficiency of methods to automatically label rhetorical roles of sentences in legal documents. Deep learning algorithms can be efficiently and effectively used in Indian legal documents, which are less explored and it will yield superior results compared to existing works.

ISSN: 2456 - 3935
International Journal of Advances in Computer and Electronics Engineering
Volume 5, Issue 6, June 2020, pp. 1–9

REFERENCES

[1] Louise Mailhot. “Decisions, Decisions: a handbook for judicial writing”. Editions Yvon Blais, Québec, Canada, 1998.

[2] H. Cunningham, D. Maynard, K. Bontcheva, V. Tablan, “GATE: A Framework and Graphical Development Environment for Robust NLP Tools and Applications,” in Proc. of the 40th Anniversary Meeting of the ACL, Philadelphia,July 2002

[3] B. Hachey and C. Grover, “A rhetorical status classifier for legal text summarisation,” in In Proceedings of the ACL-2004 Text Summarization Branches Out Workshop., 2004

[4] Atefeh Farzindar and Guy Lapalme, ‘LetSum, an automatic Legal Text Summarizing system’ in T. Gordon (ed.), Legal Knowledge and Information Systems. Jurix: The Seventeenth Annual Conference. Amsterdam: IOS Press, 2004, pp. 11-18,2004.

[5] A. Graves, S. Fernández, and J. Schmidhuber, “Bidirectional LSTM Networks for Improved Phoneme Classification and Recognition,” in Proc. Int’l Conf. on Artificial Neural Networks (ICANN), 2005.

[6] S. Mukherjee and P. Bhattacharyya, “Sentiment analysis in twitter with lightweight discourse analysis,” Proceedings of COLING 2012, pp. 1847–1864, 2012.

[7] A. Z. Wyner, W. Peters, and D. Katz, “A case study on legal case annotation,” in Proc. JURIX, 2013

[8] A. Rahangdale and A. Agrawal, “Information extraction using discourse analysis from newswires,” International Journal of Information Technology Convergence and Services, vol. 4, no. 3, p. 21, 2014.

[9] P. Jansen, M. Surdeanu, and P. Clark, “Discourse complements lexical semantics for non-factoid answer reranking,” in Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), vol. 1, pp. 977–986, 2014.

[10] M.Saravanan, Dr. B. Ravindran and Dr. S. Raman,” Automatic Identification of Rhetorical Roles using Conditional Random Fields for Legal Document Summarization”, in Legal Knowledge and Information System, JURIX, The Nineteenth Annual Conference, Paris, IOS Press,2014

[11] J. Pennington, R. Socher, and C. D. Manning, “Glove: Global vectors for word representation,” in Empirical Methods in Natural Language Processing (EMNLP), pp. 1532–1543,2014

[12] Y. Kim, “Convolutional neural networks for sentence classification,” in Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP). Doha, Qatar: Association for Computational Linguistics, pp.1746–1751, October 2014

[13] Z. Huang, W. Xu, and K. Yu, “Bidirectional lstm-crf models for sequence tagging,” arXiv preprint arXiv:1508.01991, 2015

[14] I. Nejadghoii, R. Bougueng, and S. Witherspoon, “A semi-supervised training method for semantic search of legal facts in canadian immigration cases,” in Proc. JURIX, 2017.

[15] O. Shulayeva, A. Siddharthan, and A. Z. Wyner, “Recognizing cited facts and principles in legal judge-ments,” Artificial Intelligence and Law, vol. 25, no. 1, pp. 107–126, 2017.

[16] S. Misawa, M. Taniguchi, Y. Miura, and T. Ohkuma, “Character- based bidirectional lstm-crf with words and characters for japanese named entity recognition,” in Proceedings of the First Workshop on Subword and Character Level Models in NLP, pp. 97–102,2017

[17] N. M. Schrimpf, “Using rhetorical topics for automatic summarization,” Proceedings of the Society for Computation in Linguistics (SCiL), pp. 125–135, 2018.

[18] P. Bhattacharya, K. Hiware, S. Rajgaria, N. Pochhi, K. Ghosh, and S. Ghosh, “A comparative study of summarization algorithms applied to legal case judgments,” in Proc. ECIR, 2019.

[19] Vu D Tran, Minh L. Nguyen, Kiyoaki Shirai, Ken Satoh,”An Approach of Rhetorical Status Recognition for Judgments in Court Documents using Deep Learning Models”,ieee explore,2019

[20] Paheli Bhattacharya, Shounak Paul, Kripabandhu Ghosh,S aptarshi Ghosh, Adam Wyner, “Identification of Rhetorical Roles of Sentences in Indian Legal Judgments,” in 32nd International Conference on Legal Knowledge and Information Systems (JURIX) 2019

[21] Vern R. Walker, Krishnan Pillaipakkamnatt, Alexandra M. Davidson, Marysa Linares, and Domenick J. Pesce ,”Automatic Classification of Rhetorical Roles for Sentences: Comparing Rule-Based Scripts with Machine Learning”, , in the Proceedings of the Third Workshop on Automated Semantic Analysis of Information in Legal Text (ASAIL 2019), as part of the 17th International Conference on Artificial Intelligence and Law (ICAIL 2019), Montreal, Canada, June 21, 2019.

Authors Biography

Divya Mohan, is a Research Scholar in Division of Computer Science and Engineering, School of Engineering, Cochin University of Science and Technology. She is a B.Tech. and M.Tech. holder in Computer Science. Her research interests are artificial intelligence, and machine learning.

Dr. Latha Ravindran Nair, is working as Associate Professor in Division of Computer Science and Engineering, School of Engineering, Cochin University of Science and Technology. She is a B. Tech, M.Tech and Ph.D holder in Computer Science. She has published several papers in the areas of machine intelligence and natural language processing. She has extensive contributions in Malayalam language computing. Her areas of interests are machine intelligence, natural language processing and image processing.

Cite this paper:

Divya Mohan, Latha Ravindran Nair, "A Survey on the Effective Machine Learning Approaches for Rhetorical Role Labeling of Sentences in Legal Documents", International Journal of Advances in Computer and Electronics Engineering, Vol. 5, No. 6, pp. 1-9, June 2020.