IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-6, Issue-5, May 2018 E-ISSN: 2347-2693

Classification of Legal Judgement Summary using Conditional Random Field Algorithm

S. Santhana Megala

Dept. of Computer Technology, SNMV College of Arts & Science, Coimbatore, Tamil Nadu, India

**Corresponding Author: bsanthanamegala@gmail.com*

Available online at: www.ijcseonline.org

Accepted: 20/May/2018, Published: 31/May/2018

Abstract— An Automatic Summary generation process creates a shortened version of the text using a Digital programming Technology, with the aim of holding the most advanced important points of the original text. In a Common Law system, previous judgments were referred to the current case arguments as well as decision making. Thus there is a need to view the previous judgments and to grasp and analyze the important points present in the legal judgments. Text Summarization technique helps the legal experts to read the key points present in a judgment just by reading the Head note generated by the system. Such techniques save the time as well as the manpower. In this paper, an automatic Legal Judgment Summarization system was implemented and tested by Fuzzy Logic, Classification and Segmentation techniques among that based on the experimental study Fuzzy Logic and Conditional Random Field Algorithm produces a meaningful summary.

Keywords—Classification, CRF, LDA, Fuzzy Logic, Legal Judgement

I. Introduction

Data Mining, as well as knowledge discovery, is the system assisted process of creating by removal through and analyzing massive sets of data and then extracting the useful information present in the data. It is also a method of discovering insightful, interesting, and novel patterns, as well as descriptive, understandable, and predictive models from large-scale data [1]. In a simple way, Data mining refers to the process of extracting knowledge that is of interest to the user.

Due to the drastic increase of digital information in the web, technology permits the system to perform the summarization process to access the shortened version of the digital information [2]. Such technology was implemented in various fields to enhance the progress of work carried out related to that.

Nowadays Legal Experts need the research community to do some technological invention to minimize their work pressure and to speed up the process. Thus the summarization method was implemented in the legal field, to enhance the judgment summarization process.

Indian Legal System follows the Statutes as well as the Common Law. Statutes were the legislative process or regulations issued by the Government, while Common Law was developed by the judges through decisions of courts and tribunals. In detail, a Common law is also called as ‘Precedent', a rule of law which is established by a court for the first time for a particular type of case and after that it is referred for decision making in similar cases.

Decisions of the judges are the sources of law. At the present time, legal professionals were carrying out the complex clerical work of interpreting the legal points and summarizing the previous judgment contents for their case arguments or to make the decision from them, Such process needs accuracy and speed. Human-generated summaries need more time and manpower and are relatively expensive.

Generating the judgment summary is a tedious task also. Thus NLP based Summarization Techniques fulfill the needs of the legal experts in a simple and efficient manner.

This research paper is organized as follows, Section I contains the introduction of Data Mining and the Importance of NLP, Section II contain the Research Methodology of the Judgement Summarization and Classification System, Section III contains the process of summary extraction using Fuzzy Logic, Section IV contains the architecture and essential steps for Classification using Conditional Random Field and Latent Dirichlet Allocation algorithms section V describes the Experimental Results and Datasets used for Evaluation, Section VI concludes research work with results arrived.

II. RESEARCH METHODOLOGY

In previous researches, Probabilistic and the rule-based techniques were used for the text summarization, but Legal judgment summarization is a tedious process, and it is not easy to find out the important sentences like any other document. A single word that occurs only one time in the judgment may belong to an important one. Hence to obtain a good Judgment summarization, enhanced methods were needed.

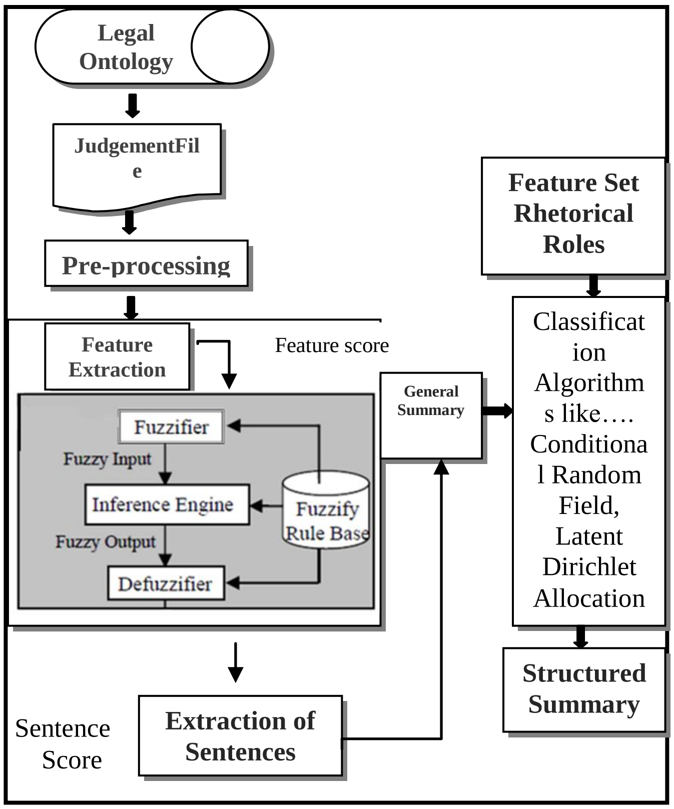

Figure 1 – Overall System Architecture of a Legal Judgement Summarization System

In this paper, efficient methods like Fuzzy Logic, Latent Dirichlet Allocation, and Conditional Random Field Algorithms were used to produce a Legal Judgment Summarization System. Overall Architecture of the 'Legal Judgment Summarization System' was depicted in Figure 1.

A Legal Ontology was created to provide a user-friendly environment. A Framework was designed to get the queries from the Legal experts and based on the data given by the legal persons, the system will display the past judgments. From that, a particular judgment file is sent to the preprocessing phase.

In the preprocessing phase, a stemming algorithm "TWIG" was proposed, which was based on the porters stemming algorithm. A slight modification and certain rules were implemented to avoid the errors and to produce a meaningful stem. Fuzzy Logic was used to summarize the legal judgment by identifying the important sentence present in that. It is based upon the values of the 11 feature extraction techniques that used in this paper.

An Advanced segmentation technique is needed to identify the structure of the legal document for that, CRF& LDA was used. CRF i.e., Conditional Random Field algorithm and Latent Dirichlet Allocation are the best algorithms used for the text segmentation based on the rhetorical roles present in the legal judgment.

III. SUMMARIZATION USING FUZZY LOGIC

Fuzzy logic is a multi-valued logic which is similar to human thinking and interpretation. It has the potential of combining human heuristics into computer-assisted decision making.

Fuzzy logic is a rational logical system, which is an expansion of multivalued logic. The fundamental concept followed by the Fuzzy Logic is that of a linguistic variable, i.e, a variable whose values are words rather than numbers [8]. In other words, Fuzzy logic is not like a digital software which recognizes only binary functions or real values like 4.6, 7.1. Instead, it is similar to a human thinking and interpretation, and it gives meaning to expressions like "High", "Small" and "Medium".

Fuzzy Logic is described as a superset of Boolean logic (i.e., '0' or '1') that has been extended to handle the concept of partial truth (i.e., Values in between '0'to '1'). Thus, Fuzzy Logic deals with "Degrees of Truth". According to [9] one major Advantage of using Fuzzy Logic is that it can deal with real-world vagueness.

One of the hard-hitting jobs in any fuzzy Logic based software application is presently how to translate observed inputs into a fuzzy membership value, and subsequently, further create the rules leading the use of connectives such as AND and OR for the fuzzy set.

A.Fuzzy Sets

A Fuzzy Set consists of a linguistic variable i.e., the values are words but not numerical. On the other hand, a classical set contains "CRISP" boundaries. The transition from one value to another is gradual and each value will determine a membership function which represents the degree to which it belongs to that value.

B.Membership Function:

A *Membership Function*, which is denoted as MF, is a curve that describes how each and every point in the input space is mapped to a membership value (or degree of membership) between 0 and 1. The input space is sometimes referred to as the *Universe of Discourse*, X.

C. Fuzzy Rules:

The fuzzy rule is based on "if…then" rule and connects the different input and output fuzzy variables. It can be expressed as

If x is A and y is B then z is C

where A &B are the antecedents and C is the consequent. Fuzzy rules are similar to common sense rules as they resemble human thinking and are based on human experience.

D. Fuzzy Inference System

Fuzzy inference system (FIS) is a framework which is based on fuzzy sets, fuzzy rules, and fuzzy reasoning. It has four main components including fuzzifier, rule base, inference engine and defuzzifier.

The fuzzifier creates fuzzy sets like “Very Low, Low, Medium, High and Very High” from “crisp” values obtained from the Feature Extraction techniques and the values for the Member Function will be divided into “Unimportant, Average and Important”.

A. APPLYING FUZZY LOGIC TECHNIQUE

Summarization process has been implemented using Fuzzy Logic Technique, which involves four steps, namely, (1) Initialization, (2) Fuzzification, (3) Inference and (4) Defuzzification.

Algorithm: the Fuzzy Logic algorithm for Summarization

Step 1: Initialization

- Defining the linguistic variables and terms // which are sentences or words from a natural language, instead of numerical values
- Construct the membership functions //which are used to quantify a linguistic Term
- Construct the rule base // used to control the output variable, which contains simple IF-THEN rules)

Step 2: Fuzzification

- Convert crisp input data to fuzzy values using the membership functions

Step 3: Inference

- Evaluate the rules in the rule base
- Combine the results of each rule

Step 4: Defuzzification

- Convert the Fuzzy output data to crisp output data using the Centroid Method

Fuzzy logic extorts the important sentences by means of Fuzzy Rules and Membership Functions based on their sentence features.

The four main components of the Fuzzy Logic System are Fuzzifier, Inference Engine, Defuzzifier and Fuzzy Knowledge Base.

In the Fuzzifier section, the membership function translates the inputs into linguistic values, such that the values from ‘0’ to ‘1’ was divided as “Very low”, “Low”, “Medium”, “High” and “Very High”.

The inference engine refers to the Fuzzy rule base, which contains Fuzzy IF-THEN rule to derive the linguistic values. At last, the Defuzzifier converts the linguistic variables to the final crisp values from the inference engine, using output membership function. The final sentence score was derived.

In the Defuzzification step, the output membership function step is divided into three membership functions, namely, "Unimportant", "Average" and "Important", which convert the result of the inference engine into a crisp output to obtain a final sentence score for each sentence.

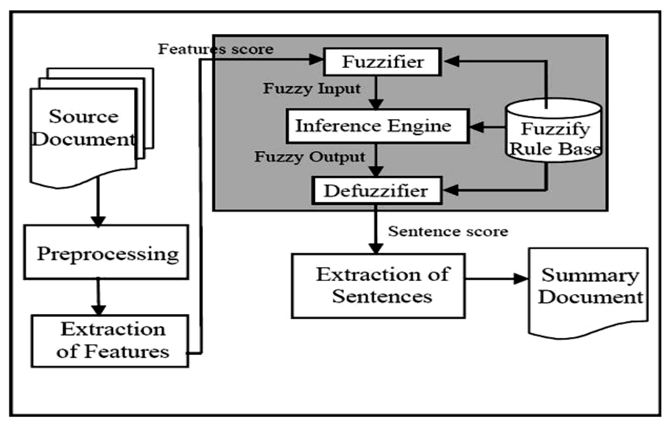

Figure: 2 – Architecture of the Text Summarization System based on Fuzzy Logic

The Membership Functions used in the Fuzzy Logic was based on the Fuzzy Centroid Method, which calculates the score for the entire sentences present in the legal document. Fuzzy Centroid Method used generalized triangular membership function to obtain the sentence score, which depends on the three parameters ‘a', ‘b' and ‘c', in which the position of the parameters 'a' and ‘c' are left and rightmost feet of a triangle and ‘b' is the peak of a triangle. The output value was obtained from zero to one for each sentence, based on the sentence features and knowledge base. The above-said value shows the degree of importance of the sentences present in the final summary. The formula to calculate the fuzzy centroid (1) is given below.

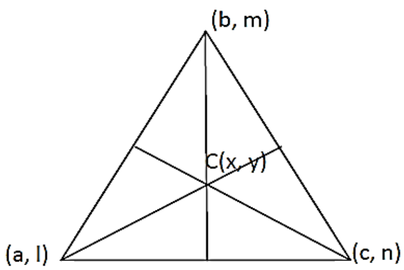

Figure 3 - Centroid Calculation.

$$C(x,y) = \left(\frac{a+b+c}{3}, \frac{l+m+n}{3}\right) \quad (1)$$

The values a, b and c were the standard values of Low, Medium and High respectively and the values l, m and n were the calculated values of Low, Medium and High respectively. Defining IF-Then rules is important in the Inference Engine. Sample IF – Then rules for the Inference Engine based on the Feature Extraction measures are mentioned below.

Table 1- Sample Fuzzy Rule

IF (Sentence Position is VH) or (Proper Noun is H) or (Sentence Length is VH) or (tf*isf is H) or (Sentence to sentence Similarity is VH) or (Citation is H) or (Local & Layout Features is VH) or (Paragraph Structures is H) or (Thematic Word is H) or (Indicators/Cue Phrases is M) or (Legal Thesaurus is H) THEN (Sentence is important)

B. SUMMARY GENERATION

Based on the fuzzy logic method described above, each sentence present in the document is represented by a sentence score such as ‘0’, ‘0.5’, ‘1’. Then, all the sentences of a document are ranked in a descending order based on their scores, such that the least scored sentences were ranked lower and the highest scored sentences were ranked Higher [4].

A set of highest score sentences are extracted as document summary based on 20% compression rate, if all the High score i.e., “1” should be included in the summary. If it is not the case, then the compression rate will increase to include all the highest values into the summary. Thus the summary generation process may repeat for some time to generate the summary, depending upon various compression rates. Finally, the sentences present in the summary are arranged in the original order.

IV. CLASSIFICATION OF STRUCTURED SUMMARY

The most significant task in this paper is to identify the rhetorical roles in the Legal Judgement. It is a part of selection analysis, which is to be carried out to understand the meaningful textual contents present in the passage. Generally, a document is segmented into comprehensible paragraphs known as rhetorical roles. Text segmentation problem concentrates on how to identify the role boundary, where one region of text ends and another begins, within a document. Legal judgments are complex in nature and it is difficult to track the existence of different topics (rhetorical schemes). Automatic segmentation of legal text focuses on the identification of key roles, so that they may then be used as the basis of the arrangement of sentences at the time of final summary generation. In this paper, a set of training documents in the legal domain is used to train the text segmentation algorithm for the purpose of improving the role identification results. Here two text segmentation techniques were taken into studies such as CRF and LDA.

A. CONDITIONAL RANDOM FIELD MODEL

Conditional Random Fields, which is denoted as CRFs, is a probabilistic framework, used for labeling or tagging and segmenting structured data, such as sequence data, trees, and lattices. The underlying idea is that of defining a conditional probability distribution over label sequences given a particular observations sequence, rather than a joint distribution over both label and observation sequences.

The most important benefits of CRFs over Hidden Markov Models i.e, HMMs is their conditional nature, ensuing in the relaxation of the independence assumptions required by HMMs in order to make a certain tractable inference. The architecture of the CRF approach, to generate a summary for the given legal judgment document was depicted in Figure 4.

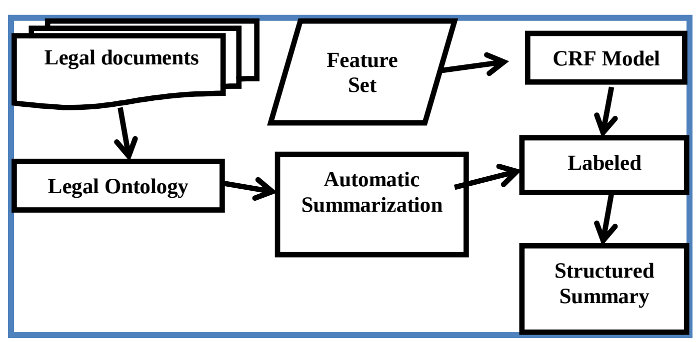

Figure 4 - Architecture of generating structured Judgment Summary using CRF

Algorithm: Construction of topic Segment

Input:
A training sentence x.
Output:
A set of segment candidates S.
Process:

```
Obtain the segment units U=(U1, U2,......Um) by
preorder traversal of the parse tree T, each Ui
    corresponds to a node in T.
For i=1 to m do
        J←i-1
         While j<m -1 and common group
         (Ui,.....Uj+1) do
        J←j+1
endfor
For k=i to j do
        For t=0 to j-k do
                s← segment(Uk,.....Uk+t)
        endfor
S←S U s
endfor
Return S
```

CRFs make a first-order Markov independence assumption with binary feature functions to link the output nodes of the graphical model in a linear chain by edges and thus can be understood as conditionally-trained finite state machines (FSMs) which are suitable for segmentation and sentence labeling [10].

A linear chain CRF with parameters C = {C1,C2,…..} defines the conditional probability for a label sequence L = l1, l2,…..,lw given an observed input sequence

$$P_C(L|S) = \frac{1}{Z_s} exp\left[\sum_{t=1}^{W} \sum_a C_a F_a (l_{t-1}, l_t, S)\right] \quad (2)$$

where Z_s is the normalization factor that makes the probability of all state sequences sum up to one, f_a (l_{t-1}, l_t, s) is a feature function which is generally binary valued and C_a is a learned weight associated with the a^{th} feature function. For example, a feature may have the value of 0 in most cases, but given the text "Points for Consideration", it has the value 1 along the transition where l_{t-1} corresponds to a state with the label Identifying the case, l_t corresponds to a state with the label History of the case, and f_a is the feature function PHRASE= "Points for Consideration" belongs to s in the sequence. To be precise, a rhetorical role identification problem in a legal domain needs to define the binary feature in the form of

$$v(S_t) = \begin{cases} 1 & \text{if the sentence contains the word "act"} \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

Now define each feature function fa as a pair a = (v, l), where v is a binary feature of the observation st and lt is a destination state:

$$f_{(v,1)}(l_t, S_t) = \begin{cases} 1 & \text{if } v(S_t) = 1 \text{ and } l_t = 1 \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

Similarly, define feature function for transitions between different label states l and l' as follows:

$$f_{(l',1)}(l_{t-1}, l_t) = \begin{cases} 1 & \text{if } l_{t-1} = l' \text{ and } l_t = l \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

Large positive values for Ca indicate a preference for such an event, while large negative values make the event unlikely and near zero for relatively uninformative features. These weights are set to maximize the conditional log-likelihood of the labeled sequence for a training set D = {(s_t, l_t): t = 1,2,…N}, written as

$$L_C(D) = \sum_i log\, P_C(L_i|S_i)$$

$$= \sum_i \left(\sum_{t=1}^{W} \sum_a C_a F_a (l_{t-1}, l_t, S) - log\, Zs_i\right) \quad (6)$$

The training state sequences are fully labeled and definite, the objective function is convex, and thus the model is guaranteed to find the optimal weight settings in terms of L_C(D). The most probable label sequence for an input sequence s_i can be efficiently calculated by dynamic programming using modified Viterbi algorithm [5].

VERTIBI ALGORITHM

The Viterbi algorithm [5] provides an efficient way for finding the most likely state sequence, in the maximum a posteriori probability sense of a process assumed to be a finite-state discrete-time Markov process.

The Viterbi algorithm has an awful constructive property of providing the best interpretation, which gives the entire context of the observations present in the data.

The working principle of Viterbi Algorithm is, to find the most likely path through a trellis, i.e. shortest path, given a set of observations.

The trellis, in this case, represents a graph of a finite set of states from a Finite States Machine (FSM).

Each node in this graph represents a state and each edge possible transitions between two states at consecutive discrete time intervals

Here a Conditional Random Field (CRF) with state space S were given, initial probabilities πi of being in state i and transition probabilities ai,j of transitioning from the state 'i' to state 'j' and outputs y1, y2,……, yT.

The most likely state sequence X1, X2, ……., XTthat produces the observations is given by the recurrence relations [6] given below:

$$V_{1,k} = P(y1 \mid k).\ \pi_k \quad (7)$$

$$V_{t,k} = max_{x \in S}(P(y_t|k).\ a_{x,k}.V_{t-1,x}) \quad (8)$$

In equation 8 $V_{t,k}$ is the probability of the most likely state sequence dependable for the first 't' observations that have 'k' as its final state.

The Viterbi path can be retrieved by saving back pointers that remember which state x was used in the second equation.

Let Ptr(k,t) be the function that returns the value of x used to compute $V_{t,k}$ if t>1, or k if t=1. Then:

$$x_T = argmax_{x \in S}(V_{T,x}) \quad (9)$$

$$x_{t-1} = Ptr(x_t, t) \quad (10)$$

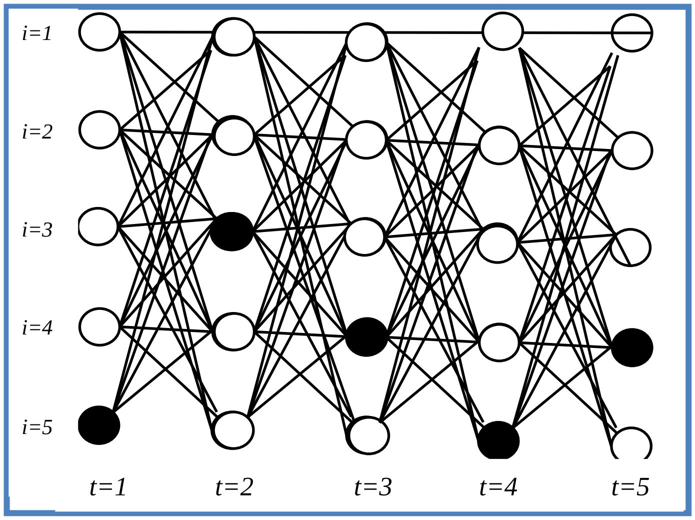

Figure 5 - Keep tracking the likely sequence states using VA

The Viterbi algorithm used to finding the shortest route through a graph is shown below:

Algorithm : Viterbi

// Input

The observation spate O={o_1,o_2,....o_N}
State space S={s_1,s_2,....s_K)
Sequence of Observations Y={y_1,y_2,.......y_T)
Transition Matrix A_{ij} of size K·K
Emission Matrix B_{ij} of size K·N
An array of Initial Probabilities π

//Output

X={x_1,x_2,......x_T} // the most likely hidden state sequence

// Process

```
Function VITERBI (0,S,π,Y,A,B) :X
    For each state Si do
        T1[i,1]←πi.Biy1
        T2[i,1]←0
End for
For i←2,3,....T do
    For each state sj do
        T1[j,i] ←Maxk(T1[k,i-1 . Akj .Bjyi)
        T2[j,i] ←argmaxk(T1[k,i-1 . Akj .Bjyi)
    End for
End for
ZT←argmaxk(T1[k,T])
XT←SZT
    For i←T,T-1,......2 do
        Zi-1←T2[Zi,i]
        Xi-1←SZi-1
    End for
Return X
End function
```

B. *LATENT DIRICHLET ALLOCATION*

Latent Dirichlet Allocation, which is denoted as LDA, is a generative probabilistic model for collections of distinctive data such as text documents [7]. The data present in the documents are symbolized as a finite fusion over an underlying set of topics which, in turn, are representation of an infinite mixture over a fundamental set of word probabilities.

Thus the topics provide an explicit symbol of the documents. Topics for the given documents can be obtained using a simple algorithm in Natural Language Processing. The architecture of the LDA approach, to generate summary for the given legal judgment document was depicted in the Figure 6.Based on the following parameters, a vocabulary of W distinct words, a number of topics K, two smoothing parameters α and β, and a prior distribution over document lengths, LDA creates a random document whose contents are a mixture of topics.

Based on the use of LDA, the set of sentences present in documents were divided into topics. LDA uses a Dirichlet distribution to represent these topics and under the Dirichlet distribution, these variables are independent of each other.

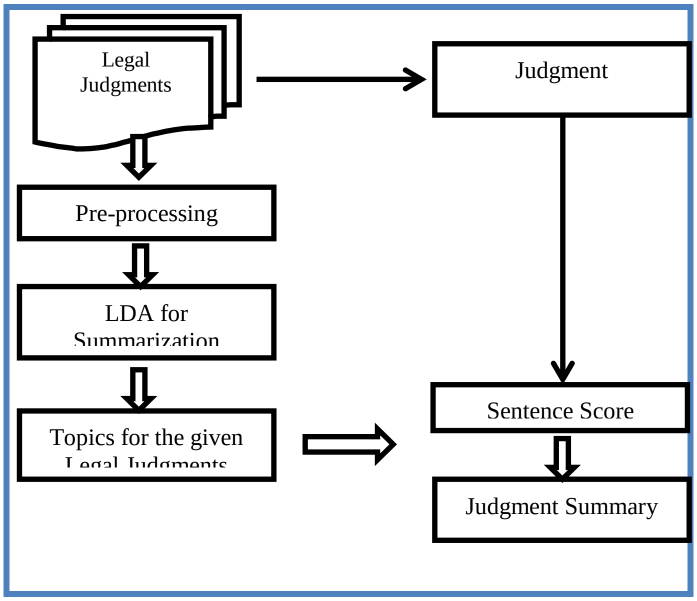

Figure 6 - Architecture to generate Legal Judgment summary using LDA

After the pre-processing all the sentences present in the document were sent to LDA as bag of words and the outcome of the process is some different topics, based on the probabilistic model. Now a set of topics for the given corpus using LDA topic model is derived. Consider each judgment from the corpus and find the sentences present in the document using Sentence Boundary method.

The algorithm to find the sentence score for each sentence from the given judgment is shown below. Consider all the sentences S_r , r Є {1,...,R} in the documents and all the Topics T_j , j Є {1,...,K} and then by calculating the probability of the Sentence S_r for the given the Topic T_j i.e. $P(S_r |T_j)$. Thus calculating the probability for the sentence S_r belongs or represents the topic T_j . Let the words of the sentence S_r be $\{W_1,W_2, \ldots W_q\}$. Algorithm to find sentence score for each sentence based on each topic for the entire corpus was given below:

Algorithm : Sentence Score Generation

Input:
D= $\{d_1, d_2,\ldots,d_m\}$ //Documents in the corpus for summarization
$T_j=\{ T_1, T_2,\ldots,T_j \}$ //Topics from LDA
Output:
S= $\{s_1, s_2,\ldots,s_m\}$ // Sentence score for each sentence for each document
Process
for each document d_i D do T_j
for each Topic T_j do
for each sentence S_r d_i do
$P(S_r |T_j) = P(W_i|T_j)*P(T_j|D_m)$
end for
endfor
endfor

Now the sentence score for each sentence based on the topic was calculated, the next step is to find the summary, consisting of maximum of two sentences from each topic. The algorithm to find the summary is shown below:

Algorithm : Judgment summary using LDA

Input:
D= $\{d_1, d_2,\ldots,d_m\}$ //Documents in the corpus for summarization
$T_j=\{ T_1, T_2,\ldots,T_j \}$ //Topics from LDA
Output:
Summary= $\{s_{m1}, s_{m2},\ldots,s_{mm}\}$ // Summary of the each document in the set
Process
For each document d_i € D do
For each Topic T_j do
S_i=Sentence_Score(d_i, T_j)
Arrange the sentences in the descending order based on the sentence score
End for
For each topic T_j do
Select top 2 sentences from S_{mij} whose score is greater than or equal to average score considering all the sentence in that document
If (any of the sentences appears already with respect to the previous topic) then
Select the next sentence.
End for
Arrange the sentences according to the sentence number of d_i to S_{mi}
End for

The algorithm takes each document from the given corpus as input and sentence score for each sentence is calculated using Sentence Score with respect each topic. The top 2 sentences with respect to each topic are selected for the final summary by eliminating redundancy. The final summary is obtained for each legal judgment by arranging the extracted sentences according to the sentence number in the original document.

- STRUCTURED SUMMARY:

A rhetorical role can be classified based on the type of the case or based on the law or based on the user. Hence, in this paper, the legal judgment is classified based on the eight rhetorical roles, which is mentioned in Table 2.

Table 2 - The description of the rhetorical role present in the Proposed research work

Rhetorical Role	Description
Headnote	Name of the Court, Judge, Date, Petitioner & Respondent Name
Identifying the case	Identify the issues to be decided for a case, it is also called as “Framing the issues”.
Establishing facts of the case	The facts that are relevant to the present proceedings/trials.
History of the case	The sequence of events with factual details that led to the present case.
Arguments	The court discussion based on the arguments made by the advocates with reference to the statute and precedents
Ratio decidendi	The reason for the application of any legal principle/law to decide a case Judgment. It is also described as "central generic reference of a case".
Final decision	It is a final decision or conclusion of the court
Reference / Citation	Previous case References used in the Legal Judgment

A rich set of features were included in this paper to identify the rhetorical roles present in the Legal judgment.

V. EXPERIMENTAL RESULTS

A number of preliminary experiments were conducted to test the efficiency of the proposed system. This chapter explains those experiments carried out using the designed methodology. The sample data sets used for the experiments, the techniques used to evaluate were discussed below. Results of the experiments along with the metrics were also discussed along with the related assumptions and hypotheses.

A. Data Sets

In this paper, a sample input of 150 legal judgments in the field of Service Law, Law of Torts and Constitutional Law was used to carry out the experiments. Among these 150 Legal Judgements, 50 Judgements were from Service Law, 50 judgments from Law of Torts and 50 documents from Constitutional Law.

These legal judgments were collected as a digital version (Soft copy) from the legal websites, The Judgement Information System (http://judis.nic.in/) and Indian Kanoon (http://indiankanoon.org/). The actual copy of the Judgement from the court is available in the above-said websites, these text documents were downloaded and utilized in our proposed system for the experimental purpose.

B. Performance Evaluation

Evaluating the quality of a summary is a very ambitious task. Many critical questions were outstretched when concerning the suitable methods and the type of Evaluation [3]. Multiple ranges of sources were available to compare the performance of the summarization system. Thus a text summary generated by a system can be compared to a human annotated summary, or to the summary generated by the online summarizers or to the summary generated by other methods. In this paper, the following metrics were used to evaluate the performance of the system such as Precision, Recall, F- Measure, Receiver Operating Characteristic Curve, and t- Test.

- PRECISION, RECALL, AND F-MEASURE

Precision and Recall are the basic measures, which is used in evaluating the text summaries. Precision is a measure of the accuracy on condition that a specific thing has been predicted. The recall is a measure of the capacity of a prediction model to select occurrences of a certain thing from a data set. A single measure that trades off precision versus recall is the F-Measure, sometimes called as F-Score, is a weighted Harmonic Mean of the Precision and Recall.The Precision, Recall, and F-Measure were calculated for the seven segments based on the rhetorical roles, which show a good ratio between them. The Average F-Measure Value shows a good reliability. The sample results of precision, recall, and F- Measure are shown for a single legal document

$$Precision\ (P) = \frac{No.\ of\ relevant\ Sentences,\ retrieved}{Total\ Number\ of\ Sentences\ retrived} => \frac{N_O}{N_m} \tag{11}$$

$$Recall\ (R) = \frac{No.\ of\ relevant\ Sentences\ retrieved}{Total\ Number\ of\ relevant\ Sentences} => \frac{N_O}{N_h} \tag{12}$$

$$F\text{ - }Measure = \frac{2\ PR}{P\ +\ R} \tag{13}$$

Table: 3- Precision, Recall & F- Measure Value for the Seven Segments using CRF& LDA

SEGMENTS IN JUDGMENT 1	PRECISION P= (No/Nm)		RECALL R=(No/Nh)		FMEASURE F=2((P*R)/(P+R))	
	CRF	LDA	CRF	LDA	CRF	LDA
Identifying the Case	1	0.9	1	0.8	1	0.85
Establishing the Facts of the Case	0.6	0.4	1	0.85	0.75	0.54
History of the Case	0.78	0.5	0.95	0.81	0.85	0.62
Analysis	0.28	0.1	0.25	0.19	0.26	0.13
Ratio Decidendi	0.73	0.42	0.85	0.73	0.78	0.53
Final Decision	0.54	0.49	1	0.8	0.70	0.61
Reference / Citation	0.86	0.72	0.8	0.73	0.82	0.72
	Average F -Measure				0.737	0.572

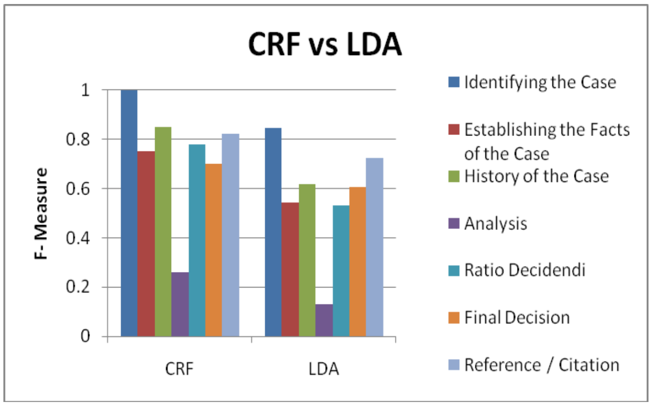

Figure: 7 - Graphical representations of F-Measure value for CRF & LDA

- PAIRED T-TEST

A Statistical Paired T-Test, normally called as T-Test, is used to compare and evaluate two population means, wherein observations in one sample can be paired with observations in the other sample.

In this paper, the average F-Measures of the sample Legal Documents were taken as the performance measures for the statistical test. A Null Hypothesis H0 was set by stating that there is no difference between the results generated by the Conditional Random Field & Latent Dirichlet allocation. On the other hand, an Alternative Hypothesis H1 indicating that there is a difference between the results generated by the Conditional Random Field & Latent Dirichlet allocation. Based on the Statistical Paired t-test results denoted in Table 4, it clearly states that CRF method provides results better than the LDA Method.

Table 4 – T- Test table for the Null Hypothesis H0 based on the average F-Measures obtained for CRF & LDA.

Documents	Fuzzy Logic		Online Summarizer	
	True	False	True	False
Judgment 1	0.26	0.01	0.06	0.03
Judgment 2	0.57	0.1	0.3	0.19
Judgment 3	0.8	0.3	0.59	0.45
Judgment 4	0.89	0.58	0.7	0.68
Judgment 5	0.91	0.79	0.73	0.89

Based on the above results it clearly shows that the Null Hypothesis that stated was rejected because t value calculated is greater than the t critical value ie., 7.039 > 2.306. Hence the p-value obtained for the calculated t score is 0.00054, which is less than 0.01, ie., 0.00054<0.01. Therefore the result is significant at p<0.01. The hypothesis testing was framed with the significance level 0.01 at a confidence level 99%.

- RECEIVER OPERATING CHARACTERISTIC CURVE

The ROC curve is a fundamental tool for diagnostic test evaluation. It shows the tradeoff between the sensitivity and specificity. The ROC is also known as a Relative Operating Characteristic Curve because it is a comparison of two operating characteristics (TPR and FPR) as the criterion changes.

The true positive rate is the ratio of the number of relevant sentences retrieved to the total number of relevant sentences in the document. The false positive rate is the ratio of a number of irrelevant sentences retrieved to the total number of irrelevant sentences in the document.

$$TPR = \frac{No.\ of\ relevant\ Sentences\ retrieved}{Total\ Number\ of\ relevant\ Sentences} => \frac{N_0}{N_h} \quad 14)$$

$$FPR = \frac{No.\ of\ irrelevant\ Sentences\ retrieved}{Total\ Number\ of\ irrelevant\ Sentences} => \frac{N_i}{N_n} \quad (15)$$

Table: 5 - Comparison between online summarizer and fuzzy logic based on TP & FP Rate

Methods	n	Mean	SD	T value calculated	T Value Critical	Df	P	Decision
CRF	9	0.229	0.010	7.039	2.306	8	0.000054	Reject
LDA	9							

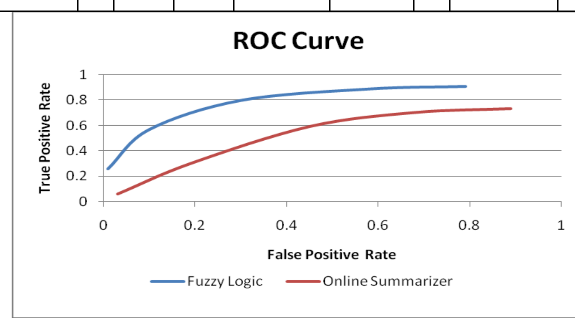

Figure: 8- ROC Curves for Fuzzy Logic & Online Summarizer

ROC analysis is related in a direct and natural way to cost/benefit analysis of diagnostic decision making. ROC Curve for our system grows on the left top border, which shows a good accuracy.

- SAMPLE OUTPUT

The sample output for the Legal Judgment file Oriental Insurance Company Ltd. vs. Kausalya Devi held at 2011 was shown in Figure 9, it is a general summary generated by the Fuzzy Logic technique. Then the structured summary for that same Judgment was shown in Figure 10 which is generated by using the Conditional Random Field

Algorithm based on the Rhetorical Roles present in the Legal Judgment document.

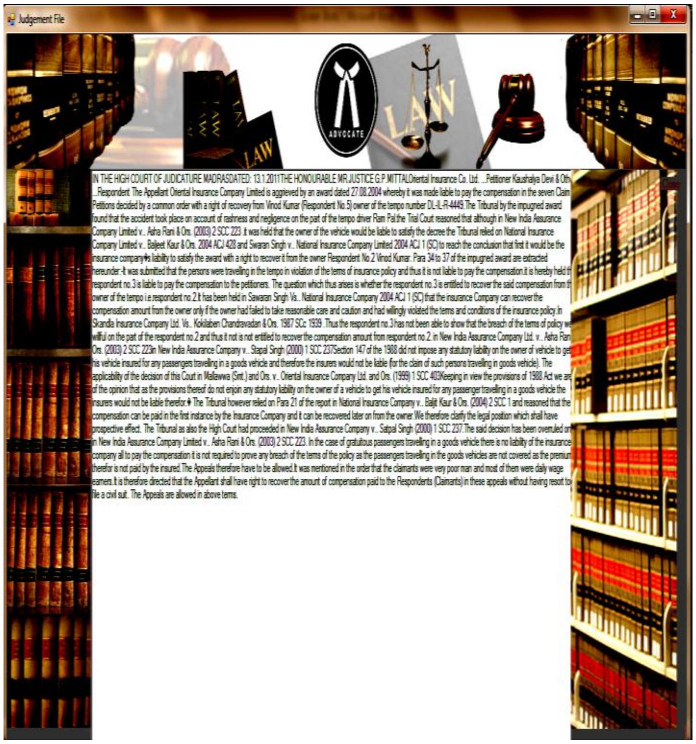

Figure: 9 – Sample Output for Judgement Summarization using Fuzzy Logic

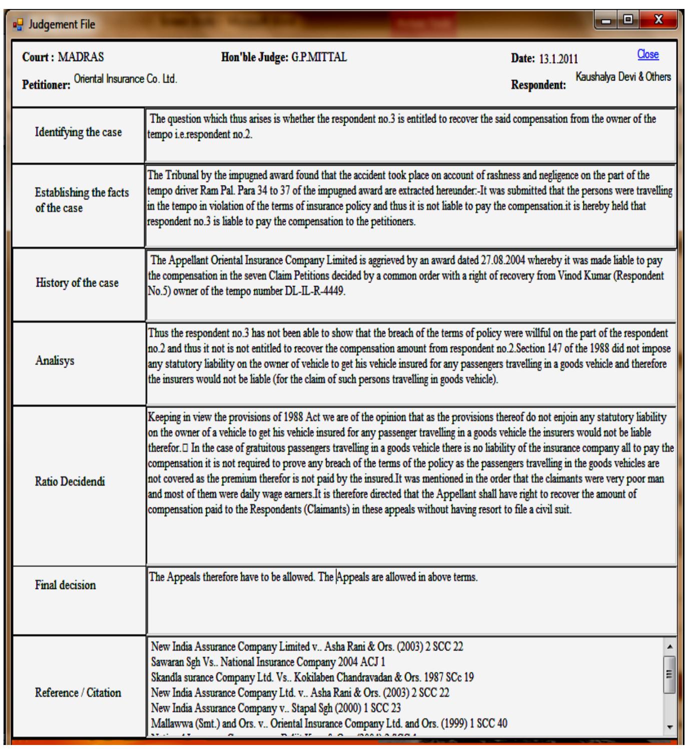

Figure: 10 – Sample Output for the Structured Summary using Classification Technique

VI. CONCLUSION

This Research work was implemented in three sub domains namely, Service Law, Constitutional Law, and Law of Torts. Next, the System was tested and evaluated with the publicly available online automatic Summarization tools and it is better than the compared one. The Experimental results of this research paper show a commendable result for Conditional Random Field algorithm while comparing with Latent Dirichlet Allocation algorithm. Thus the combination of Fuzzy Logic and CRF gives more accuracy as well as it produces meaning full summary and it segments the topic accurately while comparing with the Latent Dirichlet Allocation algorithm.

REFERENCES

[1] Mohammed Zaki J, Wagner Meira, "Data Mining and Analysis: Fundamental Concepts and Algorithms", Cambridge University Press, 2014.

[2] Mani I, Maybury M, "Advances in Automatic Text Summarization", Cambridge MIT Press, 1999.

[3] Shams R, Elsayed A, and Akter Q.M, "A Corpus-based evaluation of a domain-specific text to knowledge mapping prototype", Special Issue of Journal of Computers, Academy Publisher, 2010.

[4] Patil M.S, "A Hybrid Approach for Extractive Document Summarization Using Machine Learning and Clustering Technique", (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (2), PP: 1584-1586, 2014.

[5] Viterbi A.J, "Error Bounds for Convolution Codes and Asymptotically Optimal Decoding Algorithm", IEEE Transactions on Information Processing, vol. 13, pp. 260-269, 1967.

[6] Ryan M.S and Nudd G.R, "The Viterbi Algorithm" Warwick Research report 238, Department of Computer Science, University of Warwick, Coventry, England, Feb.1993.

[7] Nenkova, Lucy Vanderwende, and Kathleen McKeown. "A compositional context sensitive multi-document summarizer: exploring the factors that influence summarization". In SIGIR 2006, New York, NY, USA, ACM, PP: 573-580, 2006.

[8] Kyoomarsi F, Khosravi H, Eslami E and Davoudi M, "Extraction-Based Text Summarization using Fuzzy Analysis", Iranian Journal of Fuzzy Systems Vol. 7, No. 3, pp. 15-32, 2010.

[9] Ladda Suanmali, Naomie Salim, and Mohammed Salem Binwahlan, "Fuzzy Logic Based Method for Improving Text Summarization", (IJCSIS) International Journal of Computer Science and Information Security, Vol. 2, No. 1, 2009.

[10] Ravi Kumar V and Raghuveer K, "Legal Documents Clustering using Latent Dirichlet Allocation", International Journal of Applied Information Systems, 2012, P: 34-37

[11] Pramod Pardeshi, Ujwala Patil, "Fuzzy Association Rule Mining-A Survey", International Journal of Scientific Research in Computer Science and Engineering, Vol.5, Issue.6, pp.13-18, 2017.

[12] A. Yadav, V.K. Harit, "Fault Identification in Sub-Station by Using Neuro-Fuzzy Technique", International Journal of Scientific Research in Computer Science and Engineering, Vol.4, Issue.6, pp.1-7, 2016

[13] Amit Palve, Rohini D.Sonawane, Amol D. Potgantwar, "*Sentiment Analysis of Twitter Streaming Data for Recommendation using,*

Apache Spark", International Journal of Scientific Research in Network Security and Communication, Vol.5, Issue.3, pp.99-103, 2017.

Authors Profile

Dr.S.Santhana Megala pursed pursued Bachelor of Science from Madurai Kamaraj University in the year 2006 and Master of Computer Applications from Madurai Kamaraj University in the year 2009. She completed her Ph.D. in PRIST University in the year 2017 and currently working as Assistant Professor in Department of Computer Technology, SNMV College of Arts & Science, Coimbatore She has published more than 7 research papers in reputed international journals. Her main research work focuses on Data Mining, IoT and Big Data Analytics and Cloud Computing. She has 8 years of Teaching Experience.

