Computer Science Review 40 (2021) 100388

Contents lists available at ScienceDirect

Computer Science Review

journal homepage: www.elsevier.com/locate/cosrev

Review article

Summarization of legal documents: Where are we now and the way forward

Deepali Jain *, Malaya Dutta Borah, Anupam Biswas

Department of Computer Science and Engineering, National Institute of Technology Silchar, Assam, 788010, India

ARTICLE INFO

Article history:
Received 27 May 2020
Received in revised form 28 January 2021
Accepted 18 February 2021
Available online 9 March 2021

Keywords:
Text summarization
Legal document summarization
Legal summarization tools
Legal domain

ABSTRACT

Due to huge amount of legal information availability on the internet, as well as other sources, it is important for the research community to do more extensive research on the area of legal text processing, which can help us make sense out of the vast amount of available data. This information growth has compelled the requirement to develop systems that can help legal professionals as well as ordinary citizens get relevant legal information with very little effort. In this survey paper, different text summarization techniques are surveyed, with a specific focus on legal document summarization, as this is one of the most important areas in the legal field, which can help with the quick understanding of legal documents. This paper starts with the general introduction to text summarization, following which various legal text summarization techniques are discussed. Various available tools are also described in this paper which is used for summarization of legal text. Two case studies are also presented in this work, where the automatic summarization of heterogeneous legal documents from two countries is considered. With the presented detailed review of the state of the art approaches, comparative analysis from the case studies and also discussions on several important research questions, this work is expected to provide a good starting point for researchers to perform a more in-depth exploration of the area of legal document summarization, more specifically with respect to the key future research directions identified in this work.

Contents

1. Introduction

This age of data deluge has resulted into the growth of online information very rapidly each day. This sort of online information growth is also seen in law field in the form of legal documents [1].

* Corresponding author.
E-mail addresses: deepali_rs@cse.nits.ac.in (D. Jain), malayaduttaborah@cse.nits.ac.in (M.D. Borah), anupam@cse.nits.ac.in (A. Biswas).

https://doi.org/10.1016/j.cosrev.2021.100388
1574-0137/

A document is legal if the intention behind creating is enforcement in the court of law. A legal document is also called as 'written performatives' by Austin [2]. These documents include constitutions, contracts, deed, orders/judgements/decrees, pleadings, statutes, wills. These documents are quite elaborative in terms of structure from a general document and are very long to read and understand [3]. It would be better if shorter versions are available for these long documents in the form of summaries. Summaries are the shorter versions of long documents which includes all relevant information.

The main motivation behind carrying out the extensive literature review of legal document summarization, is that across the world, legal information is produced in large amount by the numerous legal institutions. In India itself, there are 25 High Courts [4] and 672 District Courts [5] which publish the legal reports publicly. This is of supreme important because several cases are pending in Indian courts [as of 2019, 87.5 percent, District and Subordinate courts] [6]. Since legal notes are long documents, so, legal institutions engage legal experts to produce headnotes which is known as summary. But, this is remarkably time-consuming task as it requires extensive human participation. Thus, automatic summarization of legal documents can significantly help legal practitioners, thereby also reducing human efforts significantly [3]. With the use of automatic text summarization techniques, legal document headnotes (summaries) can be generated.

One of the ways to get such automatic summaries is to use automatic text summarization techniques, which can produce summaries without losing the relevant information of the document under consideration [7]. Such kind of automatic summarization technique has very high utility in the field of law, which has led to the introduction of Automatic Legal Document Summarization Domain—a sub-domain of text summarization in general [8].

Currently, the generation of legal document summary is a process where considerable amount of human effort is involved. This process is labour-intensive, time-taking and expensive. For example, legal professionals like lawyers and judges need to send the cases to legal experts for creating summaries for them. Apart from this, these legal professionals need to refer to previous similar cases in order to prepare their own defences as well as provide verdicts [8]. The novice readers of legal documents also want to get an idea on a current case as well as previous related cases, without having to go through a huge number of complex legal documents [9]. Now-a-days, legal documents are often very easily available through online sources [10–12], so that ordinary citizens can also access them. Automatic summarization tools are also very helpful for ordinary citizens because using such a system, summaries of any case can easily be accessed. This also leads to a very high degree of transparency [13], since such kinds of tools help get rid of a lot of hard to understand legal jargon. Thus automatic summarization tools can prove to be of very high utility, in the field of law, thereby facilitating fast processing of legal cases by legal professionals, quick understanding of past cases by all the stakeholders, as well as a very high level of transparency.

The domain of automatic legal document summarization differs from text summarization in general, because these documents are often presented in many different structures, depending upon the country of origin for the case, and also the heavy usage of information carrying citations make the task of summarization even more challenging in this domain [3]. For example, consider Figs. 1(a) and 1(b) below, which shows a general structure of legal document from United States (US) and from India respectively. The two documents are very different in terms of their structure, which can introduce significant difficulties in developing a general legal document summarization tool.

Due to the peculiarities of legal documents, some key research questions arise in the field of legal document summarization, which are given below:

- **RQ1:** The legal documents from different countries vary vastly, in terms of a number of factors like document structures, lengths, etc.. How do they affect the quality of automatic summarization?
- **RQ2:** What are the metrics available to check the quality of summary? Is the evaluation metric efficient enough to always give good results?
- **RQ3:** How can the quality of legal summarization be improved by performing other upstream Natural Language Processing (NLP) tasks?
- **RQ4:** How to achieve better structuring of legal document summaries?
- **RQ5:** Why has there been a lot of work for extractive legal summarization, while less or no work for abstractive legal summarization?

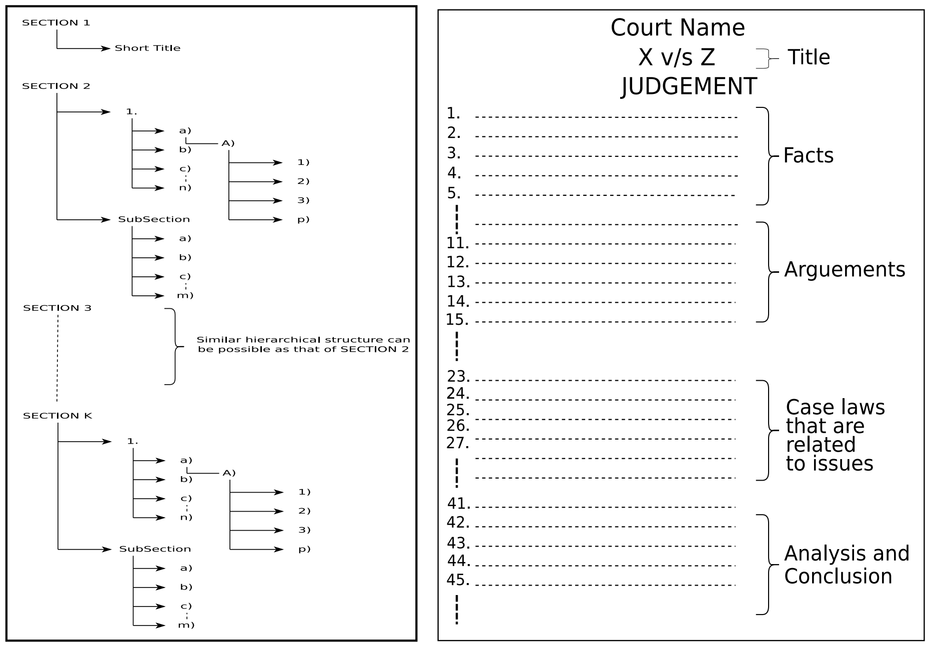
In this work, a detailed survey is conducted so that it enhances the understanding of the reader about legal document summarization, as well the reader is able to find answers to some of the most important research questions in this domain. The main contribution of this survey work can be summarized in the following points:

- In order to understand the current state of the automatic legal document summarization domain, an extensive literature survey is performed.
- Several important research questions have been identified which point towards the need of doing research in specific areas of legal document summarization.
- A comparative analysis of several country specific legal documents is performed for the task of text summarization.
- After performing comparative analysis, several key observations are drawn that help understand the current state of the techniques for summarization. Also, multiple limitations have been identified which motivate specific potential future research directions in this domain.

The paper is divided into 8 sections. Section 1 starts with the general introduction of summary, then it tells the importance of summarization in general and in legal field. Section 2 comprises of a discussion of text summarization in general, where some of the state of the art works in the area of extractive and abstractive text summarization are discussed. Evaluation metrics are discussed in Section 3. Section 4 discusses various domain independent and domain specific legal document summarization techniques. Then, in Section 5, some of the available legal document summarization tools are discussed. Two legal document summarization case studies are presented in Section 6, considering legal documents from US and India. The case studies are enriched with a detailed comparative analysis of several summarization techniques in the domain. Following which in Section 7, the findings of the literature survey are used to address the research questions identified in the introduction section. Moreover, the limitations of the current work in the domain are identified in the discussion section, with the help of which future research directions are proposed. Finally the paper is concluded in Section 8 with a summarization of the findings of the literature survey work.

2. Text summarization

The method in which shorter versions of documents are produced automatically without compromising on their actual meaning, is known as text summarization. With the growth of online



(a) Structure of a typical US legal bill (b) Structure of a typical Indian case Judgement

Fig. 1. Snippet of legal document structure from two different countries.

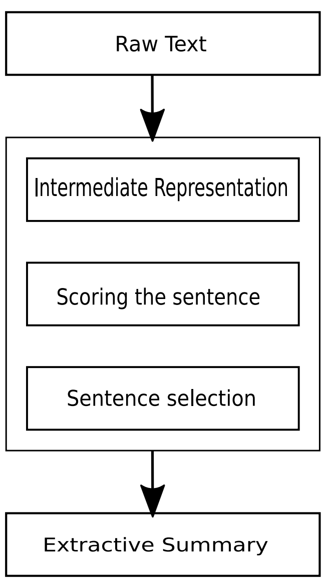


Fig. 2. Extractive summarization model.

documents which are being produced each minute of everyday, it becomes relevant to have such systems which can automatically create summaries. These summaries can be of two types: (1) Extractive and (2) Abstractive. Extractive summarization is the generation of summary containing sentence subset of the original text after identifying the important sentences; whereas abstractive summarization can be defined as summarizing a document in our own words after reading the full document, without losing the original meaning. There are several detailed literature surveys on text summarization in general [14–16].

2.1. Extractive summarization

Traditional extractive summarization approach consists of three independent tasks: (1) intermediate text representation, (2) scoring sentences with respect to text representation and (3) sentence selection. There have been several works done in the previous years on extractive summarization using different kinds of approaches. A pictorial representation of the typical steps involved in extractive summarization is depicted in Fig. 2.

- **Statistical and Semantic features:** In this approach several statistical and semantic features are considered. Statistical features, such as word frequency, cue words, location, title are used by [17–20] in order to find the important sentences and thus the summary. There have been several works, such as [21,22] which deal with both statistical and semantic features such as term frequency, word frequency, inverse sentence, stop words filtering, resolved anaphora, word senses and textual entailment for creating extractive summary.
- **Machine learning based approach:** Machine learning includes supervised and unsupervised approaches. Supervised machine learning includes those approaches in which training examples have labels. While unsupervised machine learning deals with those scenarios where training examples do not have labels. One of the approaches for performing unsupervised machine learning is clustering. In this approach, based on similarity, clusters are formed. This cluster formation could be of either sentences or document itself. Many researchers have explored this approach in the past such as [23–25]. One of the works in which authors have examined the most widely used sentence scoring methods for extractive text summarization using machine learning techniques is [26]. Selection of relevant sentences is the main aim of this work. One of the papers which deals with the formation of extractive summary [27], is based on the sentence importance classifier in which important sentences are predicted first and then summary is formed as per required length.

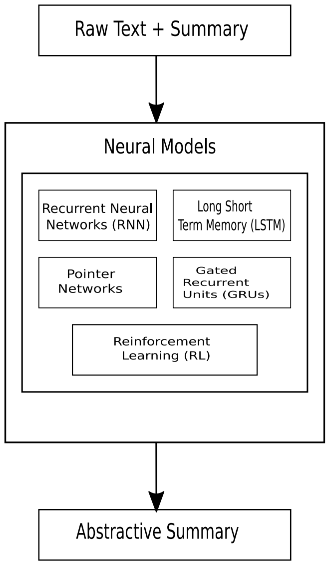


Fig. 3. Abstractive summarization model.

- **Probabilistic approaches:** The main goal of this approach is to find salient sentences, key concepts and relationships among those concepts. Various models such as the Bayesian model, Hidden Markov Model (HMM), etc., have been used for finding such important concepts through probability [28, 29].
- **Graph based approach:** In this approach, construction of a similarity graph takes place in which vertices and edges are determined using the similarity matrix of sentences. In this similarity graph, sentences are the vertices and similarity scores between sentences are given as edge weights. Then, Pagerank algorithm [30] is used to determine sentence important. Then, ranking of sentences is done based on these scores, and finally top k sentences are selected as summary. Some of the work using this approach includes [31–35].
- **Neural network based approach:** This type of approach includes learning of text through neural networks. With the recent advancements in the field of deep learning, several works have been done, such as [36–42], which are based on neural extractive summarization approaches.
- **Text simplification based:** The goal of this technique is to reduce any lexical or syntactic complexity associated with text without losing the meaningful content. This step is considered as a preprocessing step in which simplification of a sentence leads to selection of informative sentences, from which an effective summary can be formed. Some of the works which have considered this type of approach includes [43,44].

2.2. Abstractive summarization

Abstractive summarization has been thought as one of the challenging tasks for a very long time and due to this, research mainly revolved around extractive summarization. But with the emergence of deep learning models, abstractive summarization does not remain a difficult task and today most of the research revolves around abstractive summarization. Fig. 3 shows the abstractive summarization model. There have been many works done for abstractive summarization, with respect to various approaches:

- **Encoder–Decoder based approach:** In this approach, input is read through encoder and encoded into a fixed length internal representation which is then used by decoder for decoding until the last sequence of text is reached. Several works in the literature have considered this approach [45–49].
- **Pointer generator network:** The concept of pointers [50] is used which helps in copying words from source text [51]. This network is a kind of balance between extractive and abstractive summarization.
- **Reinforcement Learning:** The idea of reinforcement learning is also explored for training the encoder–decoder based summarization models and to tackle the problem of repetition [52,53].
- **Hierarchical model:** Recently multiple text summarization approaches have been proposed, that focus on a hierarchical approach, for achieving efficient summarization. For example, the authors in [54], have used hierarchical structure for sentence representation. Whereas [55] have used sentence classification layer over summarization layer in order to derive a hierarchical structure, where the labels from sentiment classification is further treated as 'summarization' of the text summarization.

Apart from these more common approaches, recently another type of abstractive summarization system has been proposed by [56], that creates a more informative summary by applying rule based text generation methods.

3. Evaluation metrics

The performance of an automatic summarization approach is tested through multiple evaluation metrics, among which ROUGE scores and it's variants are by far the most popular. ROUGE [57] stands for Recall-Oriented Understudy for Gisting Evaluation. The number of overlapping units such as word pairs, n-gram and word sequences between system generated summary and human generated reference summary is automatically counted by the ROUGE score. There are four different variant of rouge score which includes ROUGE-SU, ROUGE-W, ROUGE-N, ROUGE-L and ROUGE-S which are described below:

- **ROUGE-N:** It measures the n-gram overlapping between candidate generated summary and human generated summary. It is computed by the formula given in Eq. (1).

$$ROUGE-N = \frac{\sum_{S \in ReferenceSummaries} \sum_{gram_n \in S} Count_{match}(gram_n)}{\sum_{S \in ReferenceSummaries} \sum_{gram_n \in S} Count(gram_n)} \tag{1}$$

  where S is a sentence, $gram_n$ and $count_{match}$ is the matching of n-grams between system generated summary and a set of reference summaries. Here, n is n-gram's length.
  ROUGE-1 is a unigram. ROUGE-2 is a bi-gram. ROUGE-3 is a tri-gram and ROUGE-4 is a 4-gram. The n-gram number increases with increase in reference summary in denominator. Therefore, this metrics integrate multiple references. The numerator part comprises of summation of all reference summaries. Thus, in this way, it gives more weightage to those n-grams which matches in reference summaries. Therefore, ROUGE-N favors those candidate summaries that share more words in reference summaries.
- **ROUGE-L:** The quality of the system summary is captured by measuring the longer Longest Common Subsequence (LCS) between the system and the human summaries. The two summaries are said to be similar when there is the longest LCS between candidate and reference summaries. Inclusion

of longest in-sequence common n-grams is done automatically by this metric and hence, this metric does not require to define n-grams before matching. This metric somehow, overcomes the limitations of ROUGE-N metric, to be more precise, the fact that ROUGE-N metric measures the similarity based on shorter sequences of text, ROUGE-L considers the LCS between the two sequences of the text. This metric is more sophisticated than ROUGE-N, but still, it suffers from the fact that there should be continuous n-grams.

- **ROUGE-W:** In this metric, weight is given to the longest common subsequence. One of the problems with the ROUGE-L score is that it gives same score for the sentences which have same LCS. The LCS can either be a consecutive long sequence of words or long sequence with gaps. It is not wrong in saying that long sequences with gaps should receive lower score. ROUGE-W has extra weighting function which gives more weight to consecutive longer sequences than longer sequences with gaps. A weighted score of 1.2 is introduced which measures the contiguous common subsequences.
- **ROUGE-S:** It is a statistics metric which measures the co-occurrence of skip-bigram. It actually measures the overlapping of skip-bigrams. It allows gaps between any pair of words in an order in a sentence, hence, its name, skip-gram.
- **ROUGE-SU:** This metric is the extension of the skip-gram metric. Extension is done by adding unigram as counting unit in ROUGE-S metric. Rouge-S has a problem of not including those candidate sentences which do not have any word pair match with its reference sentences. To deal with the problem of ROUGE-S, an extended version is introduced which is known as ROUGE-SU that also consider the unigram matching between two summaries.

Apart from the ROUGE score, the three most frequently used measures are recall, precision and F-measure, which are also widely used in the information retrieval domain.

Precision (P): Out of total number of documents retrieved, how many are actually relevant documents retrieved as defined below in Eq. (2).

$$P = \frac{\#relevant\ documents\ retrieved}{\#retrieved\ documents} \tag{2}$$

Recall (R): Out of total number of existing relevant documents, how many are actually relevant documents retrieved as defined below in Eq. (3).

$$R = \frac{\#relevant\ document\ retrieved}{\#relevant\ documents} \tag{3}$$

F-measure (popularly used F_1) measures the harmonic mean between precision and recall as defined below in Eq. (4).

$$F_1 = 2 \cdot \frac{P \cdot R}{P + R} \tag{4}$$

4. Legal document summarization

The process in which summaries are generated from legal text which includes court judgement documents, bills under process as well as acts and laws, is called legal document summarization. It is important to note that legal documents are significantly different from other types of general texts in many aspects, which makes this task of summarization very challenging. One of the earlier and seminal works in this field was done in 2004 in which Grover et al. [58] introduced a legal dataset which contains 188 judgements from House of Lords Judgement (HOLJ) website from 2001–2003 for the extractive summarization of British judgements. In the upcoming sections, some domain independent and domain specific techniques are described. But before that, there is a need of understanding the differences which makes legal documents different from the general documents. Legal documents are different from the documents in other domains in many aspects which makes it difficult to create a summary for legal documents. Turtle [1] reports various aspects in which legal documents are different.

- **Legal document size:** The length of legal documents are comparatively longer than the documents in other domain. It is because the other documents still depend upon abstracts rather than the whole documents.
- **Structure of legal document:** Legal documents generally have a wide range of internal structure. The internal structure follows a hierarchy, that is often country specific.
- **Vocabulary:** Legal documents follow their own legal-based terminology besides the standard language.
- **Ambiguity:** The ambiguity in legal documents lies in the fact that, these documents contain terms, phrases that may have different meanings with respect to different legal institutions.
- **Citations:** The Citation plays an important role in legal documents. It is the major issue of legal documents because they contain citations that cannot be ignored.

4.1. Domain independent unsupervised extractive text summarization algorithm

There are some classical algorithms which can be applied to any domain. Some of these algorithms such as Lexrank, Latent Semantic Analysis (LSA) and Textrank, which have also been widely applied in the legal domain, are described below:

- **LexRank:** In order to determine the sentence importance, Erkan and Radev [34] presents a stochastic graph based approach. Text summarization uses a sentence salience for determining the importance of sentences. The proposed work is based on the eigen vector centrality. In this approach, intra-sentence cosine similarity based connectivity matrix is formed. The authors have also discussed various approaches to determine lexical centrality in multi-document summarization using similarity graph. The important sentences of a document is found using the concept of sentence centrality and then using those important sentences, summary is created. The dataset used was DUC 2004 and DUC 2003. The developed approach has been compared with MEAD and the results show that it outperformed centroid based summarization techniques in most of the cases.
- **Latent Semantic Analysis (LSA):** This is yet another algorithm which is widely used in NLP for performing various tasks. Landauer et al. [59] first proposed the LSA algorithm. This technique analyzes distributional semantics between the terms and their corresponding set of documents. The basic assumption here is that similar texts will contain related terms. A matrix is formed out of a large piece of text and this matrix contains words count per document. Unique words are represented by rows and each document is represented by columns. The working of LSA is based upon Single Value Decomposition (SVD) technique. The number of rows in the matrix in this technique is reduced using similarity structure is preserved among columns. Then, cosine similarity is used to check the similarity between documents. Documents are similar, if the value close to 1, while dissimilar documents have value close to 0.

  Merchant and Pande [60] uses this approach in their work. In their work, authors have proposed two summarization models for (a) untrained single documents and (b) trained multi-document. However, the effectiveness of proposed system is not totally captured by the ROUGE score.

- **Textrank:** The graph based algorithm is proposed by Mihalcea and Tarau [33]. As per similarity matrix, a graph is formed in which vertices represent sentences and edges represent edges. Then, sentence importance is found out using the Pagerank algorithm [30] based on scores. Then, top-k sentences are selected to form an extractive summary. Authors have used DUC 2002 for finding important sentences in their experiments and the results have been found out that proposed approach outperforms the baseline approaches.
  A modified Textrank [31] is proposed by authors in which edge weights are given by an inverse sentence frequency modified cosine (*isf*-modified-cosine) similarity score which finds the similarity between two sentences.
  More recently in [61], the authors have proposed the application of a fine-tuned textrank algorithm with the help of Bayesian Optimization (BO) for summarizing legal bills and the results have shown that proper fine-tuning is actually helpful to get the most out of the existing algorithms.
- **Reduction:** The main idea of yet another graph based algorithm is to remove some extraneous phrases from the sentences and thus form a summary [62]. The whole reduction process is based on the 'Graph Reduction' idea. While reducing the graph, some context information is considered. Lexical links determine the importance of words and then scores of such words is calculated. Finally, score of the sentence is determined by calculating the score of the individual words. The score tells the importance of a sentence in its local context.
- **Luhn:** HP Luhn [17] introduced a heuristic approach of summarizing the documents based on word frequency and position of that word.
  In the recent paper [63], a task of summarizing plain English contracts is done on which unsupervised approaches such as Textrank, KLSum, Lead1, etc., are applied. These algorithms have not been able to perform well due to the level of abstraction and style differences.
  In another recent work [64], a comparative analysis of several extractive and abstractive approaches is performed on the legal rulings. The results show that abstractive approaches performs significantly better than extractive approaches.
- **Restricted Boltzmann Machines (RBM)**: Verma and Nidhi [65] introduced an another unsupervised approach using deep learning technique known as RBM for performing the summarization task. This approach consists of three phases in which firstly extraction of features is done followed by feature enhancement and finally summary is generated. The authors have chosen 9 sentence features for extraction and then these features gets enhanced using RBM using a single hidden layer and 9 perceptrons with a learning rate of 0.1.

4.2. Legal document specific extractive summarization algorithms

Based on different approaches that have been followed for summarization of legal documents, a detailed investigation of various related works have been presented below:

- **Citation based:** This type of summarization depends upon the set of citation sentences to form a summary which is known as citation based summarization. In this method, some catchphrases are created which are nothing but a form of legal summary.
  Galgani and Compton [66] uses both incoming and outgoing citations which are further combined with other citances (sentences about the target document). The citances, target's document original text, cited and citing cases' catchphrases are together known as citations. The effectiveness of citation-based summarization method has been shown through results. In addition to that, the proposed approach is also flexible that it can be applied to other domains as well.
  In another work, Galgani et al. [67] also uses citation-based approach to create the summary of the legal documents. The authors have used incoming and outgoing citation information for summarization of legal documents. Extraction of catchphrases from text of target legal document is the main aim of this approach. The dataset contains 2816 cases from the year 2007 to 2009 which is taken from Federal Court of Australia (FCA). Extracted sentences are compared using ROUGE score with each catchphrase individually. The results show that catchphrases proved to be effective for the legal professionals who are looking for relevant precedents of any case and also of great help when browsing routinely through documents.
- **Rhetorical role based:** Rhetorical roles refer to the understanding of the semantic function of a sentence with the rhetorical role associated with that sentence. Rhetorical roles help in performing tasks such as summarization by aligning the different sentences which are associated with the rhetorical roles in the final summary generation. So, rhetorical roles act as an information which makes the final summary more readable and coherent. Also, a table-like summary can be generated which is a well structured summary under the different rhetorical roles. This can be thought of as thematic segmentation [68] also in which relevant sentences are extracted for each theme and finally a well structured summary is generated.
  Grover et al. [69] used the tenses as linguistic features which proved to be of great help for identifying the rhetorical roles of sentences in legal documents.
  This approach is explored by Saravanan et al. [70] using Conditional Random Fields (CRF), Hachey and Grover [71] have used Teufel and Moens' [72] approach for rhetorical labelling and hence to form a summary. Whereas Bhattacharya et al. [73], have used deep learning models for identification of rhetorical roles which can be used for performing downstream tasks such as summarization.
- **Ripple down rules based:** It is basically an error-driven approach. In this method, domain experts created rules without involving a knowledge engineer. Knowledge base is created with incremental refinements. The whole process is monitored by a system expert when system is in use. Whenever there is an error, the old rules which generates error gets replaced with new rules which are provided by the experts and these new rules are written back to the knowledge base. Where the failure occurs, that point is identified and replaced by new rules by the system in the correct position so that any incorrect interactions of rules are avoided. A smaller quantity of annotated data is required for creating manual rules which is in contrast to machine learning techniques which requires multiple instances to identify important features.
  Pham and Hoffmann [74] developed a system KAFTIE which is based on ripple down rules for handling natural processing tasks.
  Galgani et al. [75] used hybrid approach which consists of several summarization techniques in one rule based system.
- **Graph based:** One of the application of summarization is found in multi-role court debate. Unlike other dialogues, court debates can be lengthy. Due to this, important information often hides. In this regard, the various researchers

have recently proposed an end-to-end debate summarization model.

Duan et al. [76] proposes a technique in which assignment of the utterances in the court debate is learnt jointly by the model. The main aim of this learning is to give more focus on controversies and tells the appropriate location of the essential utterances which are needed for each controversy to assist the court judge to make legal decisions. The model extracts legal information behind the debates by using legal knowledge graph. Court debate records of around 279,494 are collected which are of civil Private Loan Disputes (PLD). Among them, the proposed model is evaluated on 5,477 court records of civil trials.

Kim et al. [77] presents their own graph based-method for legal text summarization. In their graph-based method, they have developed directed and disconnected graphs for each document. The number of representative sentences with coherency are chosen automatically for summarization. The authors have also proposed their own weighting scheme based on nodes/ edges in which selection of topic words is the major step. One of the advantage of this approach is that there is no need to provide compression rate, instead, they give sentence selection method which automatically decides compression rate. The authors have performed their experiments on HOLJ dataset. The proposed algorithm has been compared with [78] in which supervised machine learning algorithm is applied. Using the proposed algorithm, results have been presented which shows that the proposed algorithm outperforms the previous results. In order to compare with clustering algorithms, the author have used X-clustering algorithm, and applying algorithm to each cluster. The proposed algorithm beats the other clustering algorithm and it has also shown good results on comparing compression ratios with the gold summaries. However, the recall value decreases on adding sentences in the gold summary, which have not shared many words with decision sentence.

- **Nature inspired:** Nature inspired approaches are those optimization approaches which are inspired by how nature adapts to challenging circumstances. There are several optimization techniques such as, particle swarm optimization (PSO), genetic algorithm etc. There have been several works done for legal document summarization, that utilizes such optimization ideas.

  Kanapala et al. [79] presents a legal summarization problem as binary optimization problem in which fitness is determined using the weight of the individual statistical features. To deal with this binary optimization problem, a mathematical model is proposed based on gravitational search algorithm (GSA) whose working depends on gravity law [80]. Several objective functions such as length of a sentence, location of a sentence, sentence similarity and keywords in the legal document. After deriving objective function, a fitness function is derived. Experiments are performed on 1000 Supreme Court judgements containing their headnotes too. These judgements are taken from FIRE 2014 (Forum of Information Retrieval evaluation). The dataset covers a judgement period from 1950 to 1989. The proposed approach outperforms the genetic algorithm, particle swarm optimization, LSA, TextRank, MEAD, SumBasic and MS-Word based approaches. It has been shown through the results that the authors' proposed algorithm can be deployed in the practical scenario.
- **Machine learning based:** A supervised machine learning approach is proposed by Le et al. [81] for legal document summarization. In this proposed approach, important sentences are extracted from legal documents. The authors also proposed a technique of presenting an XML structured data which deals with specific issues in the legal data. Then, Naive Bayes, which is a sentence classification algorithm is applied on a surface set, content features and emphasis. And the best results are given by the combination of all three features which the author named as PRobabilistic Decision SUMmarizer (PRODSUM). The authors has compared the proposed approach with baseline methods and AustLII. Though proposed algorithm may not able to beat AustLII, but the system gives the best overall results.

  In the recent work by Zong et al. [82], authors have developed a two stage pipeline, in which firstly predictive sentences are chosen first with the help of convolutional neural networks (CNN) classifier, and then subset of sentences are selected for the final summary through maximum marginal relevance (MMR). This summary is represented in summary template.

  In one of the recent work, Tran et al. [83] proposes a catchphrase extraction system in which firstly, training of a scoring model is done using the CNN based architecture followed by the selection of catchphrases as the summary.

  In the paper Anand and Wagh [84], the authors have proposed a deep learning approach based on automatic sentence labelling approach and from the results it has been found that LSTM+Glove outperformed all other approaches. One of the advantages of authors' proposed approach is that it can be applied to any other domain.

A summary of all relevant techniques for legal summarization along with their findings are shown in Table 1.

5. Legal document summarization tools

Apart from the techniques developed for legal summaries, there are also a number of online tools that are present, which produces summaries that can help legal professionals. Table 2 summarizes the existing tools available for Legal Document Summarization.

- **CaseSummarizer:** Polsley et al. [85] proposes a tool known as CaseSummarizer. The working of this tool depends upon word frequency with some extra domain-specific knowledge. The engine is specifically built for legal people, that produces lists of entities, scalable summary text, and supplementary details like abbreviations present in the document, and a significance heat map of the entire text. A list of abbreviations provides a way through which user can match the phrases to the original text. In this way, it helps user to recognize which entities are being referenced whenever an abbreviation appears. Depending upon the relevance score, heat map colours the sentences in the document. CaseSummarizer tool follows three steps in order to produce a summary: preprocessing, scoring of sentence relevance, and domain processing. The authors have used ROUGE-N score for evaluating their algorithm where $N =$ 1,2,3,4. The authors perform their experiments on 3890 legal cases which are taken from the Federal Court of Australia (FCA) [86] which contains cases of years in between 2006 and 2009. Five randomly selected documents are taken for evaluation. The proposed tool has been compared with AutoSummarizer [87], TextSummarizer [88], SplitBrain [89] and SMMRY [90]. The results have shown that the CaseSummarizer tool performs better when compared to other expert summaries. One of the noticeable thing that author has pointed out is that, the tool was not able to beat human written summaries. This suggests that there is a room for further improvements.

Table 1
List of techniques, results and findings: Legal document summarization.

Authors	Technique	Corpus	Metrics	Results	Findings
Kanapala et al. [79]	Gravitational search algorithm (GSA)	FIRE 2014	ROUGE-1, ROUGE- 2	R1-P = 0.3385, R1-R = 0.5953,R1-F = 0.4316R2-P = 0.1417, R2-R = 0.2285, R2-F = 0.1749	GSA have been shown to beat GA, PSO, LSA, TextRank, MEAD, SumBasic, MS-Word.
Le et al. [81]	Chunk based approach+NLP concepts	Japanese National Pension Act, news articles from Mainichi Shimbun newspapers	Precision, Recall, F1-score	Precision=39.85, Recall = 92.79,F1 score = 47.16	Extracted keywords can be further used for summarization purposes.
Duan et al. [76]	End-to-end debate summarization model	Court records of civil trial	ROUGE-L, ROUGE-2, ROUGE-1	ROUGE-1 = 58.8, ROUGE-2 = 45.2, ROUGE-L= 64.0	(1) One of the issues of the proposed approach is to get legal knowledge. (2) The main aim is to solve controversy focuses.
Merchant and Pande [60]	Latent Semantic Analysis(LSA)	Indian Legal Judgement	ROUGE-1, ROUGE-2, ROUGE-L	ROUGE-1 = 0.58, ROUGE-2 = 0.15, ROUGE-3 = 0.35	Improvement has been done on proposed approach with untrained and multidocument trained LSA approach
Galgani and Compton [66]	Citation-based method + catchphrases	AustLII corpus	AVG ROUGE-1	R1-P = 0.655, R1-R = 0.686, R1-F1 = 0.631	(1) The authors use bidirectional citation. (2) The new cases can also be summarized which have not been cited yet.
Kim et al. [77]	Graph-based method	HOLJ corpus	Recall, F-score, Precision	Recall=36.4%, Precision= 31.3%, F-score=33.7%	The main aim is to produce a more coherent summary.
Galgani et al. [75]	Hybrid approach+Knowledge Acquisition(KA)	AustLII corpus	ROUGE-1	ROUGE-1 P = 0.690, ROUGE-1-R = 0.265, ROUGE-1-F = 0.363	Different kinds of formation are considered for knowledge acquisition at the sentence level, document level and at the collection level.
Farzindar & Lapalme [68]	(1) Linguistic based approaches (2) Thematic segmentation	Legal records from federal CANLII	ROUGE-L, ROUGE-4, ROUGE-3,ROUGE-2, ROUGE-1	ROUGE-L = 0.4518, ROUGE-4 = 0.1503, ROUGE-3 = 0.2071,ROUGE-2 = 0.3138, ROUGE-1 = 0.5750	The main aim is to create tabular structured summary by identifying relevant themes.
Saravanan et al.[70]	CRF	Civil Court Judgement	macro-averaged Precision, Recall, *F*-measure	P = 0.896 ,R = 0.864, F = 0.879	Use CRF technique to find rhetorical labels and hence to improve the structure of summary.
Hachey & Grover [71]	C4.5 decision trees, Naive Bayes, Winnow algorithm, Support Vector Machines	HOLJ	micro-averaged F-score	C4.5 = 65.4, NB = 51.8, Winnow = 41.4, SVM = 60.6	Rhetorical roles have been used for extracting relevant sentences for summary.
Bhattacharya et al.[73]	Hierarchical BiLSTM and hierarchical BiLSTM CRF	Indian Legal judgements	macro-averaged Precision, Recall, *F*-measure	Precision = 0.8396, Recall = 0.8098, *F*-measure= 0.8208	(1) Use of pre-trained embeddings have been shown improvements for classifying rhetorical roles using deep learning models. (2) Deep learning models perform better than hand engineered features.
Manor and Li [63]	Textrank, KLSum, Lead-1, Lead-K, Random-K	Plain English Contracts	ROUGE-1, ROUGE-2, ROUGE-L	ROUGE-1 = 24.38, ROUGE-2 = 7.52, ROUGE-L = 17.63	(1) The results show that ROUGE-score are much higher on DUC dataset as compared to legal dataset. (2) The authors points out to the need of simplification or style transfer system in the summarization pipeline.

(*continued on next page*)

- **LetSum:** Farzindar and Lapalme [68] developed a justice decision text summarization tool which is known as LetSum (Legal text Summarizer). The system determines the thematic structure of a judgement. Relevant sentences are identified for each of the themes of the segments in the decision and then the unimportant sentences are filtered out. It generates a summary into a tabular form. The system involves intrinsic and extrinsic evaluation. Intrinsic evaluation shows promising results.
- **HAUSS:** Galgani et al. [91] developed a framework to do an efficient summarization by integrating various base techniques into a single approach. Using Hybrid AUtomatic Summarization System (HAUSS), rules are created with the help of Knowledge Base (KB), that extracts catchphrase for legal

Table 1 (*continued*).

Authors	Technique	Corpus	Metrics	Results	Findings
Feijo and Moreira [64]	Luhn, Textrank, Lexrank, Sumbasic,LSA, KLSum, Random, NMT-Small, NMTMedium, Transformer, TransformerAAN	Brazilian Supreme Court	ROUGE-1, ROUGE-2, ROUGE-L	R1-F = 44.27, R1-P = 49.38, R1-R = 47.76, R2-F = 26.50, R2-P = 28.36, R2-R = 28.84, RL-F = 35.27, RL-P = 38.52, RL-R = 38.16	Although abstractive approaches outperforms extractive approaches but there is still a need of improvement in terms of redundancy.
Jain et al. [61]	Textrank with Bayesian Optimization	BillSum Dataset	ROUGE-1, ROUGE-2, ROUGE-L	ROUGE-1 = 0.404, ROUGE-2 = 0.194, ROUGE-L = 0.327	The results shows that proper hyperparameter tuning improves the existing algorithm.
Grover et al. [69]	Various linguistic tools such as NER, tense identification is utilized in this work	House of Lords judgements	Precision, Recall, *F*-measure	Precision = 83.74, Recall = 71.25, *F*-measure = 76.99	(1) Tense information is utilized in this work. (2) Statistical measures are also done to show the utility of proposed approach.
Zong et al. [82]	CNN classifier	Board of Veterans' Appeals (BVA) legal decisions	ROUGE-1, ROUGE-2	R1-0.269 (±0.171), R2-0.102 (±0.178)	While the scores suggests the applicability of their approach but qualitative analysis suggests the further improvement since most of the relevant aspects are missed.
Tran et al. [83]	CNN	Australian legal case	ROUGE-1, ROUGE-SU, ROUGE-W	R1-P-0.2311,R2-R-0.3084,R2-F- 0.2295, R-SU6-0.0685, R-SU6-0.1078, R-SU6-0.0537, R-W-1.2-P-0.1450, R-W-1.2-R-0.1363, R-W-1.2-F-0.1175	(1) The authors' approach is simple, do not depend upon any syntactic knowledge. (2) It also achieves comparable performance.
Anand and Wagh [84]	Feed forward neural networks (FFNN) and LSTM	Indian Supreme Court judgements	ROUGE-1, ROUGE-2 and ROUGE-L	R1-F = 0.436, R1-R = 0.542, R1-P = 0.376, R1-F = 0.245, R1-R = 0.283, R1-P = 0.217, RL-F = 0.382, RL-R = 0.501, RL-P = 0.335	A simple approach that does not require domain knowledge or feature crafting. The quantitative analysis of the proposed approach shows the utility of the proposed approach in legal domain and other domain as well.
Polsley et al. [85]	tf–idf approach and entity	Federal Court of Australia (FCA)	ROUGE-1, ROUGE-2, ROUGE-3, ROUGE-4, ROUGE-L	ROUGE-1 = 0.194, ROUGE-2 = 0.114, ROUGE-3 = 0.091, ROUGE-4 = 0.085, ROUGE-L = 0.061	(1) Sentence level summarization is useful for legal summaries. (2) Sentence extraction methods are indeed useful for producing summaries.

texts. These rules combine centrality, frequency, linguistic information and citation in a context-dependent way. The important sentences of any case is identified using some series of phrases or sentences which is referred to as catchphrases in legal community and sometimes it may be attached to judgements given by the judges. HAUSS Knowledge Base (KB) has scored a highest precision of 0.765 and 0.486 when a threshold of 0.5 and 0.7 respectively is given. It scores the highest recall and f-measure of 0.685 and 0.634 respectively with single condition HAUSS KB. With threshold of 0.5 and 0.7, it scores a highest recall value of 0.385 and a precision value of 0.343.

- **DecisionExpress:** A thematic segmentation based approach is utilized by "NLP Technologies" to develop the DecisionExpress tool. The working of this tool is mainly based on themes which are introduction, reasoning, context and conclusion. The system is well designed for describing the information like judge name who has signed the judgement, what kind of judgements judges pertain to, domain of the law and what is the subject of information. Hyperlinks are also provided to the summary and the original document with the help of this tool. Translation of judgements into the Canadian languages (English or French irrespective of the fact that the judgements are produced in which language) is done automatically using this tool.

Table 2
Tools for legal document summarization.

Tool name	Approach	Features
Case Summarizer	Word frequency+domain specific knowledge	Scalable summary
Hauss	Knowledge Base	Context dependent, purpose specific
LetSum	Linguistic features	Table style summary
DecisionExpress	Thematic segmentation	Bilingual information extraction, consistency, cost reduction

6. Legal document summarization case studies

In order to assess the performance of several classical as well as legal domain specific recent summarization approaches, in this

Table 3
Statistics of datasets.

	BillSum dataset						Indian dataset	
	US train		US test		CA test			
	Doc	Summary	Doc	Summary	Doc	Summary	Doc	Summary
Avg sent	62	6	62	6	46	9	516	53
Avg words	1451	196	1444	196	1695	370	15695	1300

section two case studies are presented. Both of these case studies focus on the automatic summarization of different types of legal documents, from two different countries. Firstly, the datasets utilized for experimental evaluation have been described. Following which an extensive comparative analysis of various automatic summarization techniques(both general as well as legal domain specific) is presented.

6.1. Dataset details:

For performing the comparative analysis, two datasets have been used of two different countries i.e. United States (US) and India.

- The first dataset is a publicly available legal benchmark dataset—BillSum [92] which consists of US Congressional Bills, that has been further split into 18,949 training and 3,269 testing dataset. The main motivation of introducing this dataset is to check the performance of the models developed over US Congressional bills on California state (CA) bills. Therefore, this dataset also has an additional 1,237 number of CA state bills.
  The dataset is cleaned to make it further useful for experiment purpose. Several preprocessing steps such as, normalization of certain words and sentences, remove semicolons, get rid of any whitespaces, complete sentences are formed by removing the bullet points and so finally to form a paragraph, get rid of special characters, also make sure that there must be gap between full stop and start of the new sentence. Similar kinds of cleaning steps are taken by the authors who have introduced the BillSum dataset [92].
  The structure of a typical legal bill is shown in Fig. 1(a), which depicts the highly hierarchical organization of a legal bill. Also the ROUGE score based importance of various sentence positions in the bill with respect to the reference summaries is presented in Fig. 4(a), from where it can be observed that the sentences occurring at earlier positions in the document usually are more important, as they tend to have higher ROUGE score based similarity with the reference summaries. It is important to note at this point that the sentence importance distribution for CA state bills is somewhat steeper as compared to the distribution for the US congressional bills, depicting the higher importance of initial sentences for the former.
- The second dataset utilized in this work corresponds to Indian legal case documents. Firstly, summaries are extracted from websites Lawbriefs [93], Cyber blog India [94], Law times journal [95] and then corresponding documents are extracted from Indian Kanoon [11]. During extraction, several similarity measures have been employed in order to extract the correct document corresponding to each summary. In this way, a total of 411 legal case documents of the Supreme Court of India along with their summaries are combined to form an Indian legal case judgement summarization dataset.
  After extraction, the dataset is stored in the *.txt* format. Then, preprocessing of the dataset is done. Several preprocessing steps such as, removing unwanted space, unwanted text, applying regular expressions, removal of unwanted headers and make sure all the characters are in UTF-8 encoding are done.
  The sentence position importance is depicted in Fig. 4(b), with respect to the reference summaries. A similar long tail distribution is observed for the Indian legal case dataset as well, however one important point to note here is that for Indian documents, the majority of the important sentences are found in the first 200–250 index positions in the document.

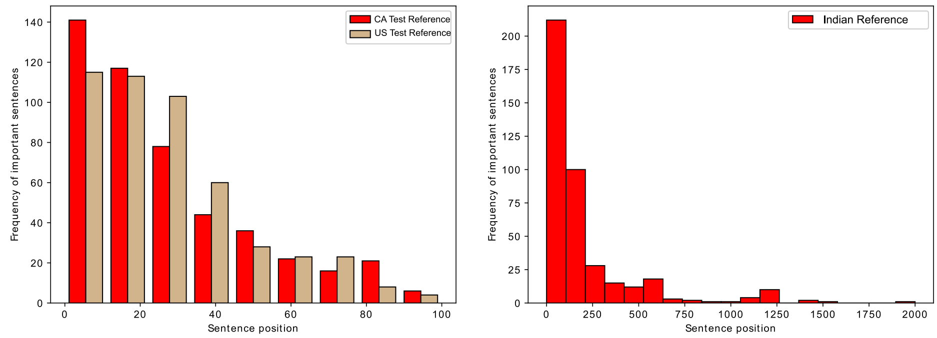
Document & Summary length: There are several other important observations for both the datasets with respect to the document and summary lengths. The average length of the documents as well as the reference summaries are mentioned in Table 3, both with respect to the number of sentences and number of words. These statistics become very important for limiting the lengths of the predicted summaries, whenever some automatic summarization technique is applied. In the case of the BillSum dataset, since the average ratio between the number of words in the train set reference summaries to the number of words in the train set documents is found to be 15%, the same ratio can be used while generating predicted summaries. Whereas, in case of the Indian legal case dataset, the summary length can be fixed to be 31%, which is the ratio of the number of words in reference summary to the number of words in the documents in the collected dataset.

6.2. Comparative analysis results

A comparative analysis is performed with the use of the algorithms that have been discussed in Section 4, with respect to the two datasets described in Section 6.1. Multiple domain independent as well as domain dependent algorithms are applied, to draw some important insights which have been discussed in this section along with the results of the comparative analysis. All the experiments were performed on a 64-bit linux machine with 16 GB RAM. A comparative analysis is done in order to verify how the techniques developed for legal documents of one country work on legal documents of another country.

The experimental results for automatic summarization in case of the US test data is depicted in Table 4. From the experimental results it can be observed that the supervised approach of LSTM based classification approach, with word2vec based representation, is the best performing model with respect to every evaluation metric except the ROUGE-1 F1, while Lexrank has achieved the highest ROUGE-1 F1 score of 0.3704.

Fig. 5 shows the sentence position importance, similar to Fig. 4(a), however both with respect to the predicted as well as reference summaries. It has been observed that important sentences are found at the starting positions, similar to the reference summaries. Moreover, the predicted summaries with the LSTM based approach has shown similar behaviour as that of US test



(a) BillSum Dataset (b) Indian Dataset

Fig. 4. Occurrence of important sentences across different positions in a reference summary.

Table 4
Performance on US test data.

Approach	ROUGE-1		ROUGE-2		ROUGE-L	
	F1	R	F1	R	F1	R
CaseSummarizer [85]	0.3402	0.3920	0.1448	0.1622	0.2851	0.2995
RBM [65]	0.2971	0.3008	0.1079	0.1131	0.2397	0.2284
LSTM with w2v [84]	0.3615	**0.6539**	**0.2086**	**0.3720**	**0.3664**	**0.5358**
LSTM with glove [84]	0.3619	0.6475	0.2071	0.3655	0.3655	0.5301
Lexrank [34]	**0.3704**	0.5415	0.1811	0.2604	0.3365	0.4230
Textrank [33]	0.3269	0.6295	0.1793	0.3423	0.3383	0.5037
LSA [60]	0.3277	0.4008	0.1288	0.1542	0.2890	0.3354
Reduction [62]	0.3472	0.5700	0.1757	0.2817	0.3304	0.4429
Luhn [17]	0.3515	0.5867	0.1800	0.2931	0.3405	0.4580

Table 5
Performance on CA test data.

Approach	ROUGE-1		ROUGE-2		ROUGE-L	
	F1	R	F1	R	F1	R
CaseSummarizer [85]	0.3632	0.3338	0.1551	0.1372	0.2947	0.2586
RBM [65]	0.3166	0.4049	0.1007	0.1350	0.2469	0.2755
LSTM with w2v [84]	0.4073	0.4638	0.1883	0.2093	0.3312	0.3588
LSTM with glove [84]	0.4071	0.4596	0.1863	0.2056	0.3322	0.3576
Lexrank [34]	**0.4144**	0.4529	0.1936	0.2083	0.3406	0.3531
Textrank [33]	0.4069	0.5055	**0.2015**	**0.2461**	**0.3457**	0.3848
LSA [60]	0.3363	0.3145	0.1313	0.1203	0.2970	0.2840
Reduction [62]	0.3996	0.4870	0.1843	0.2214	0.3255	0.3632
Luhn [17]	0.4112	**0.5112**	0.1981	0.2423	0.3447	**0.3871**

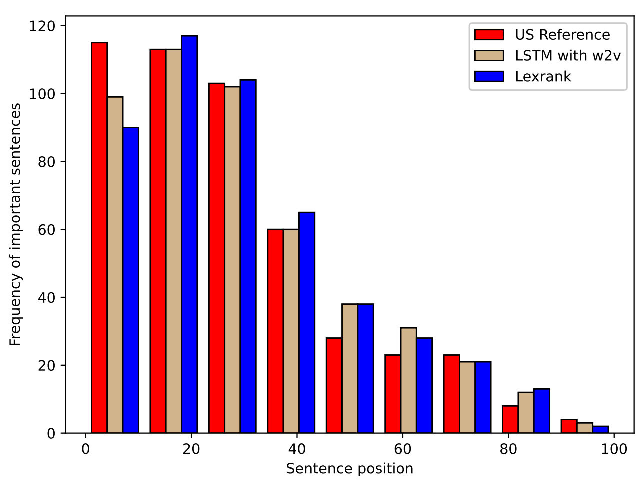


Fig. 5. Occurrence of important sentences across different positions in US Test Data.

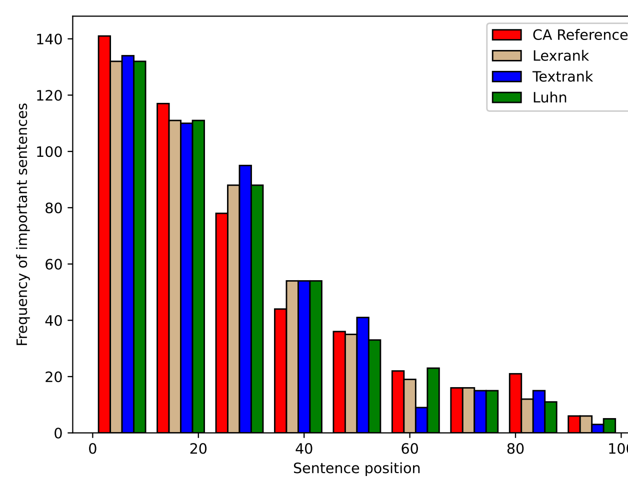


Fig. 6. Occurrence of important sentences across different positions in CA Test Data.

reference summaries, which supports the fact that it has also been the best performing model.

In case of the CA test dataset, Lexrank has achieved the highest ROUGE-1 F1 score of 0.4144 followed by Luhn, which has achieved 0.4112 ROUGE-1 F score, as shown in Table 5. Luhn has achieved the highest ROUGE-1 and ROUGE-L recall scores of 0.5112 and 0.3871 respectively. Such kind of higher recall is due to the fact that it tends to select the longer sentences from the document, to form the summaries. Whereas Textrank has been able to achieve the highest ROUGE-2 F1 and ROUGE-2 recall scores of 0.2015 and 0.2461 respectively. It has also achieved the highest ROUGE-L F1 score of 0.3457.

Fig. 6 shows a frequency distribution for the CA test dataset, depicting the positions of the important sentences in the documents, with respect to the predicted as well as reference summaries. From the figure, it is clear that most important sentences of reference summary again lie at the initial positions of the documents. The predicted summaries of Textrank and Luhn algorithms have been able to achieve similar distributions to the reference summaries, supporting their better performance with respect to the other algorithms.

In case of the Indian legal case judgement summarization task, the LSA algorithm has achieved the highest ROUGE-1 F1 score of 0.3178, as shown in Table 6. Whereas, LSTM with glove has

Table 6
Performance on Indian legal data.

Approach	ROUGE-1		ROUGE-2		ROUGE-L	
	F1	R	F1	R	F1	R
CaseSummarizer [85]	0.2750	0.6238	0.1414	0.3251	0.2767	0.4537
RBM	0.2688	0.1998	0.1120	0.0816	0.2256	0.1656
LSTM with w2v [84]	0.2889	0.6817	0.1526	0.3737	0.2783	0.5155
LSTM with glove [84]	0.2946	0.6827	**0.1551**	0.3766	0.2824	0.5179
Lexrank [34]	0.2751	0.6734	0.1440	0.3486	**0.2894**	0.4896
Textrank [33]	0.2676	0.6484	0.1341	0.3363	0.2661	0.4817
LSA [60]	**0.3178**	0.6110	0.1285	0.2550	0.2691	0.4469
Reduction [62]	0.2343	0.6712	0.1284	0.3630	0.2568	0.5027
Luhn [17]	0.2314	**0.7095**	0.1286	**0.3869**	0.2631	**0.5403**

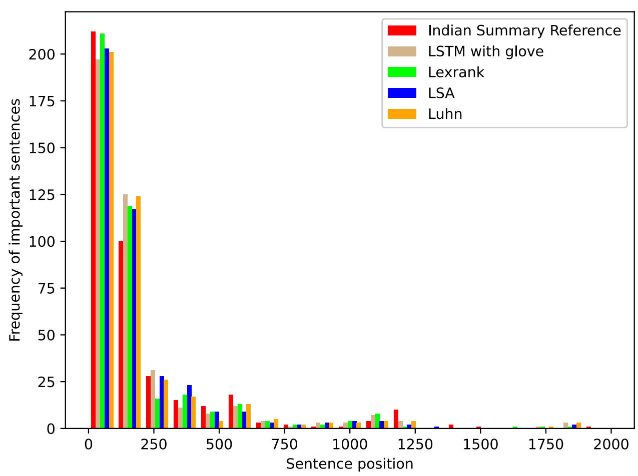


Fig. 7. Occurrence of important sentences across different positions.

achieved the highest ROUGE-2 F1 score of 0.1551. Apart from these, Lexrank and Luhn algorithms have performed quite well with respect to several metrics.

Fig. 7 shows the occurrence of important sentences in the particular document positions. It has been observed from the figure that, more important sentences are lying at the starting positions for both the reference as well as the predicted summaries. But one important point to note here is that, none of these high performing algorithms have been able to closely match the reference summaries in terms of identifying the importance sentences.

After performing the comparative analysis, several important conclusions are drawn. Firstly, from the results it has been shown that none of the algorithms can be called as the best algorithm across the datasets. The algorithm which works better for one country specific legal dataset could not perform well in case of other country specific dataset. Another important point that has been observed from the results is that supervised approaches such as LSTM with word2vec or LSTM with glove can perform better if provided with more training data and as compared to the other approaches. Moreover, the legal document specific technique CaseSummarizer, which has been applied originally on Australian legal case judgements, is not able to achieve very impressive results for any of the other country specific datasets considered in this comparative analysis. The reason being that, the performance is heavily dependent on correct identification of named entities which is difficult in case of legal documents using standard NLP tools.

7. Discussion

The research in the legal domain for summarization has its roots long back. However, a recent increase in the number of research works in this domain has been observed, mainly due to the digital availability of the legal documents, as well as improvements in automatic NLP techniques. Some of the major advantages of text summarization which may help legal practitioners such as, lawyers, judges, etc., can be summarized as follows:

- Automatic summaries help with the quick understanding of legal documents that are typically quite long in terms of number of words and sentences. This becomes even more important when the extremely long legal documents are filled with incomprehensible literary flourishes. Several newspaper articles by reputed journalists have recently questioned the understandability of such long and complex documents, citing situations where even Supreme Court judges are also not able to fully comprehend the contents of judgement documents from High Courts and Supreme Court of India [96,97]. In such scenarios automatic summarization techniques can have very high applicability.
- The summary of previous judicial decisions that are relevant for the current case, can be very helpful in making precedent based judgements. This can potentially facilitate quick processing of legal cases, which can go a long way in mitigating the typical judgement delays.
- Depending upon the type of end user, the summary need might be different. For example, a judge might be more interested in finding out judicial decision summaries, whereas a lawyer might be more interested in finding the factual summary of a legal case document [85]. Moreover, if a layman wants to understand legal documents, then the summaries need to be even simpler and it needs to cover the overall topic of the document [98]. With appropriate supervised training of computational systems on multiple types of gold standard summaries, automatic summarization systems can be developed that caters to the specific needs of the end user. Thereby, improving the overall understandability of legal documents.
- Automatic legal document summarization is of very high utility in the area of Legal Information Retrieval (LIR) as well. Legal document search engines can be enhanced with the help of automatically generated summaries of the documents, since they can be used as part of the retrieval results as snippets of the underlying documents. Moreover, the search queries for such search engines can themselves be augmented with the help of automatic summaries, resulting in better document retrieval [99].

In this survey paper, all the recent legal document summarization techniques are explored. The paper investigates several important research works that have been done in the legal document summarization area. From these studies, several important key points have been extracted out, which addresses all the research questions identified in Section 1.

In order to address **RQ1**, apart from the observations from the exploration of research works in this domain, several important observations are drawn from the case studies that have been performed in Section 6. From these observations, it is clear that, the methods that have been developed for legal documents of one country, could not perform well on the legal documents of other countries. In addition to that, several domain dependent and domain independent techniques have been utilized for performing experiments and from the results it is seen that, none of the methods can be called as the best for all the different types of documents. We can only say that comparatively, for specific scenarios a particular technique outperformed all other techniques. Such kind of variance in performance of summarization approaches can be observed from Table 1, as well as from the comparative analysis results in Tables 4–6.

One of the most popular evaluation metrics which have been used in the literature for checking the performance of summarization systems is the ROUGE metric as described in Section 3. Since this metric considers the exact matching of words, it is not always possible to find all the exact matches due to the abstractive nature of reference summaries. In the paper [60], the authors have said that the effectiveness of summarization systems cannot be completely judged by the ROUGE scores. The reason being that, the reference summaries are written by the experts and thus, there is a possibility for the words in machine generated summary that they may not match with the words in the actual/ reference summary. Similar observations are made by [85], where the authors have performed lawyer scoring of the reference as well as the predicted summaries, and significant differences between the ratings have been observed. In [3], the authors have evaluated the quality of the predicted summaries, that were generated by the best ROUGE score techniques, with the help of legal experts. This evaluation revealed that the automatic summaries were able to cover the facts and statutes properly, however they were unable to appropriately cover the precedents and holdings in legal case judgements. These observations address the second research question **RQ2**, and also motivates the need for more domain specific evaluation metrics, instead of the vanilla ROUGE score based metrics.

Several important observations have been drawn from the literature review of the current legal document summarization techniques with respect to **RQ3**. Since text summarization is a downstream NLP task, it is highly dependent on the efficient completion of comparatively simpler upstream tasks, like word vectorization, named entity recognition (NER) and sentence role classification. This becomes more evident in the legal domain, since the legal documents usually contain very specific words, named entities and also rhetorical roles of sentences. If these upstream tasks can be performed in a legal domain specific manner, improved summarization performances can be achieved. For example, if only the generic tf–idf features are used for the representation of legal texts, then this might not be able to capture all the information specific to the legal domain. Similarly, if pre-trained word embeddings are considered, then many generic embeddings are publicly available like Word2vec [100], Glove [101], etc., however development of legal specific word representations can be actually more beneficial for downstream tasks, as shown by Chalkidis and Kampas [102]. The authors have considered 123,066 documents which consists of 492 M individual words, using which Law2Vec embedddings are learnt, which show improved performance on downstream tasks. The reduced performance of generic approaches can be seen from the results of CaseSummarizer, given in Tables 4 and 5. From the results, it is can be observed that CaseSummarizer approach is not able to achieve very good performance, since the working of CaseSummarizer is heavily dependent on the named entities and with the use of generic NER tool, it is unable to recognize the legal specific words. If we consider the legal case judgements, then the need for appropriate role labelling of the sentences inside these documents, becomes very important. The availability of documents with rhetorical role labelled sentences, can enable automatic summarization approaches to perform structured thematic summarization, by forming the predicted summaries part by part where each part can cover one rhetorical role. Such kind of thematic summarization can help cover all the important parts of the legal case judgements, as shown by [68,73].

The structure of the summary, which is the main focus of **RQ4**, has been found to be of extreme importance based on the extensive literature review. It has been observed that the summary should cover all the aspects of a legal judgement document, basically the thematic or rhetorical segments which cover the entire document. For example, creation of summary using rhetorical role labelling or thematic segmentation is mainly performed in order to understand the structure of these documents. In this way, sentences are extracted from each theme/ rhetorical roles so that a more coherent and more readable summary can be produced. Thus, thematic segmentation and rhetorical labelling are the ways to achieve a well structured, readable and coherent summary.

The lack of exploration of abstractive summarization in the legal domain is quite evident from the literature review. To answer the **RQ5**, it has been observed from the literature survey, that all these studies are conducted on small legal datasets and most of the works relied on hand engineered features like cue phrases, title, sentence location, etc. It is important to note that despite many of the state of the art approaches for general text summarization being abstractive, in the legal domain, extractive summarization techniques are more commonly used. One of the reasons that can be considered is that legal documents are longer and have citations which cannot be ignored. It is possible that while using abstractive summarization, the meaning of the legal documents may get changed due to this. The second possible reason for not using abstractive techniques in the legal domain, is that abstractive techniques use deep learning approaches which require much larger datasets and most of the available legal datasets have smaller sample size. So, deep learning approaches are harder to apply on such scenarios. However, it is yet to be seen if transfer learning based ideas, which use pre-trained models just for fine-tuning for specific domains can be useful in the legal domain as well.

One of the challenges associated with the automatic summarization of legal documents is related to its fairness. If there is only one summary, it might not cover all the aspects of the legal documents. Moreover, since reference summaries are human generated and abstractive in nature, it is very much prone to high bias. To deal with such problems, multiple reference summaries need to be considered for a single document, while building benchmark datasets. Similarly, in case of the documents also, the problem of fairness becomes important due to potential under-representation of documents from different sub-domains. For example, in case of legal case judgement documents, a variety of cases should be included to make a complete dataset, along with multiple reference summaries for each such document. In such cases, one reference summary might cover all the facts and precedents while the other reference summary covers issues and arguments. So, there is a need for a heterogeneous set of documents along with multiple human written summaries, to ensure fairness and reduce human bias.

Based on the limitations that have been found through the literature review, one can consider the following as potential future work directions:

- There is a huge lack of properly labelled legal datasets. This is the biggest challenge in order to proceed with automatic legal text summarization system design and development. More specifically, legal documents cover a huge variety ranging from case judgements to patents, from legal bills to contract documents, and so on. Therefore, there is a great need for properly labelled datasets for every type of legal document, in a country specific manner. In the literature, most of the works are found on case judgement documents. Whereas, the other types are less commonly explored. Future research works can be carried out to build benchmark legal summarization datasets, that consists of a variety of legal documents, along with multiple human written summaries, so that effective automatic summarization techniques can be developed.

- One of the key observations from this literature survey work is that efficiently performed upstream tasks like NER, word representation learning, rhetorical role labelling, etc., can help improve the performance of automatic summarization of legal documents, in a significant manner. This motivates the need for future research in the development of legal specific NLP techniques. For this task also, appropriately labelled legal specific datasets need to be created, which can have a very high positive impact, not only on summarization, but also on many other complicated downstream NLP tasks in this domain.
- Since the documents in the legal domain typically are characterized by very long sentences, a simple application of extractive summarization might still result in long and complicated sentences as part of the predicted summaries. This motivates the exploration of text simplification techniques for the legal domain, which can potentially help improve the readability as well as the understandability of the automatically generated summaries of legal documents [84]. As mentioned by [63], several text simplification techniques can be explored as part of future work to further improve the quality of system generated summaries, like lexical simplification, structural simplification, unwanted qualifier removal, etc.
- As identified from the literature survey, there is a significant lack of research in the area of legal domain specific evaluation metric design, for the task of summarization. The currently used ROUGE metrics, are not able to fully capture the quality of a system generated summary. Significant future research is needed in this area, to develop a legal specific ROUGE metric, that can better capture the semantics of a summary in this domain. One potential way to achieve this, could be to ensure the inclusion of both word level, as well as phrase level semantic domain knowledge along with the traditional ROUGE metric, in order to better assess the quality of the predicted summaries. However, extensive exploration and experimental analysis is required to find such an effective measure.

Since voluminous information is generated everyday in this domain, and also there are many cases which are pending, while new cases are being added to the existing ones, it is high time to perform active research in this domain, in order to develop such methods which take into consideration all the above mentioned points and develop the techniques to deal with lengthy and complex legal documents.

8. Conclusion

In this survey paper, first an attempt is made to have a brief account, about the various text summarization techniques. The paper starts with the basic definition of text summarization and gradually describes the important techniques so that an unfamiliar reader can have better understanding about this area. Next, the main focus is shifted towards legal document summarization techniques. Legal document summarization section is started with the main issues in this domain. Then, classification of legal document summarization techniques based on different approaches like citation based, graph based, nature inspired, Latent semantic analysis approach, rhetorical roles, etc. are discussed. Various publicly available tools for legal document summarization are also described. One noticeable thing here is that only extractive work is found in the existing literature. This suggests that there is a need to do more future work in this domain.

Apart from carrying out a detailed investigation of various legal document summarization approaches, multiple key research questions have also been identified in this domain. With the help of the literature survey, these research questions have been addressed and based on the findings, several important future research directions have also been identified. Further exploration resulting in better summarization techniques, benchmark datasets and evaluation metrics for the legal domain, is found to be the key to further progress the research in this area.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding acknowledgement

Dr. Anupam Biswas' research work is funded by the Science and Engineering Research Board (SERB), India, Department of Science and Technology (DST) of the Government of India under (Grant No. EEQ/2019/000657) and (Grant No. ECR/2018/000204).

References

[1] Howard Turtle, Text retrieval in the legal world, Artif. Intell. Law 3 (1–2) (1995) 5–54.
[2] John Langshaw Austin, How to Do Things with Words, vol. 88, Oxford university press, 1975.
[3] Paheli Bhattacharya, Kaustubh Hiware, Subham Rajgaria, Nilay Pochhi, Kripabandhu Ghosh, Saptarshi Ghosh, A comparative study of summarization algorithms applied to legal case judgments, in: European Conference on Information Retrieval, Springer, 2019, pp. 413–428.
[4] Wikipedia, High courts of India — Wikipedia, the free encyclopedia, 2020, http://en.wikipedia.org/w/index.php?title=High%20courts%20of%20India&oldid=943069995 (Online Accessed 03 March 2020).
[5] Wikipedia, List of district courts in India — Wikipedia, the free encyclopedia, 2020, http://en.wikipedia.org/w/index.php?title=List%20of%20district%20courts%20in%20India&oldid=934774817 (Online Accessed 03 March 2020).
[6] Press, Press information bureau, 2020, https://pib.gov.in/newsite/PrintRelease.aspx?relid=191213 (Online Accessed 03 March 2020).
[7] Inderjeet Mani, Automatic Summarization, vol. 3, John Benjamins Publishing, 2001.
[8] Atefeh Farzindar, Guy Lapalme, Legal text summarization by exploration of the thematic structure and argumentative roles, in: Text Summarization Branches Out, 2004, pp. 27–34.
[9] Ambedkar Kanapala, Sukomal Pal, Rajendra Pamula, Text summarization from legal documents: a survey, Artif. Intell. Rev. 51 (3) (2019) 371–402.
[10] Judgement information system, 2020, http://164.100.79.153/judis/ (Online Accessed 03 March 2020).
[11] Indian Kanoon, 2020, https://indiankanoon.org/ (Online Accessed 03 March 2020).
[12] Australasian Legal Information Institute, 2020, http://www.austlii.edu.au/ (Online Accessed 03 March 2020).
[13] Michael E. Sykuta, Peter G. Klein, James Cutts, Cori K-Base: Data Overview, in: 2007 Kauffman Symposium on Entrepreneurship and Innovation Data, 2007.
[14] Mehdi Allahyari, Seyedamin Pouriyeh, Mehdi Assefi, Saeid Safaei, Elizabeth D. Trippe, Juan B. Gutierrez, Krys Kochut, Text summarization techniques: a brief survey, 2017, arXiv preprint arXiv:1707.02268.
[15] Mahak Gambhir, Vishal Gupta, Recent automatic text summarization techniques: a survey, Artif. Intell. Rev. 47 (1) (2017) 1–66.
[16] Ani Nenkova, Kathleen McKeown, A survey of text summarization techniques, in: Mining Text Data, Springer, 2012, pp. 43–76.
[17] Hans Peter Luhn, The automatic creation of literature abstracts, IBM J. Res. Develop. 2 (2) (1958) 159–165.
[18] H.P. Edmundson, New methods in automatic extracting, J. ACM 16 (2) (1969) 264–285.
[19] Kai Hong, Ani Nenkova, Improving the estimation of word importance for news multi-document summarization, in: Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics, 2014, pp. 712–721.
[20] Ani Nenkova, Lucy Vanderwende, The impact of frequency on summarization, Microsoft Research, Redmond, Washington, Tech. Rep. MSR-TR-2005, Citeseer, 2005, p. 101.

[21] Tatiana Vodolazova, Elena Lloret, Rafael Muñoz, Manuel Palomar, et al., The role of statistical and semantic features in single-document extractive summarization, 2013.

[22] Tatiana Vodolazova, Elena Lloret, Rafael Muñoz, Manuel Palomar, Extractive text summarization: can we use the same techniques for any text?, in: International Conference on Application of Natural Language To Information Systems, Springer, 2013, pp. 164–175.

[23] Ramiz M Aliguliyev, A new sentence similarity measure and sentence based extractive technique for automatic text summarization, Expert Syst. Appl. 36 (4) (2009) 7764–7772.

[24] Wenjie Li, Mingli Wu, Qin Lu, Wei Xu, Chunfa Yuan, Extractive summarization using inter-and intra-event relevance, in: Proceedings of the 21st International Conference on Computational Linguistics and the 44th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, 2006, pp. 369–376.

[25] Maofu Liu, Wenjie Li, Mingli Wu, Qin Lu, Extractive summarization based on event term clustering, in: Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics Companion Volume Proceedings of the Demo and Poster Sessions, 2007, pp. 185–188.

[26] Gabriel Silva, Rafael Ferreira, Rafael Dueire Lins, Luciano Cabral, Hilário Oliveira, Steven J. Simske, Marcelo Riss, Automatic text document summarization based on machine learning, in: Proceedings of the 2015 ACM Symposium on Document Engineering, 2015, pp. 191–194.

[27] Yinfei Yang, Forrest Sheng Bao, Ani Nenkova, Detecting (un) important content for single-document news summarization, 2017, arXiv preprint arXiv:1702.07998.

[28] Pascale Fung, Grace Ngai, Chi-Shun Cheung, Combining optimal clustering and hidden Markov models for extractive summarization, in: Proceedings of the ACL 2003 Workshop on Multilingual Summarization and Question Answering-Volume 12, Association for Computational Linguistics, 2003, pp. 21–28.

[29] Asli Celikyilmaz, Dilek Hakkani-Tür, Discovery of topically coherent sentences for extractive summarization, in: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1, Association for Computational Linguistics, 2011, pp. 491–499.

[30] Sergey Brin, Lawrence Page, The anatomy of a large-scale hypertextual web search engine, 1998.

[31] Chirantana Mallick, Ajit Kumar Das, Madhurima Dutta, Asit Kumar Das, Apurba Sarkar, Graph-based text summarization using modified textrank, in: Soft Computing in Data Analytics, Springer, 2019, pp. 137–146.

[32] Daraksha Parveen, Hans-Martin Ramsl, Michael Strube, Topical coherence for graph-based extractive summarization, in: Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, 2015, pp. 1949–1954.

[33] Rada Mihalcea, Paul Tarau, Textrank: Bringing order into text, in: Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing, 2004, pp. 404–411.

[34] Günes Erkan, Dragomir R Radev, Lexrank: Graph-based lexical centrality as salience in text summarization, J. Artif. Intell. Res. 22 (2004) 457–479.

[35] Gunes Erkan, Dragomir Radev, Lexpagerank: Prestige in multi-document text summarization, in: Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing, 2004, pp. 365–371.

[36] Yang Liu, Fine-tune BERT for extractive summarization, 2019, arXiv preprint arXiv:1903.10318.

[37] Pengjie Ren, Zhumin Chen, Zhaochun Ren, Furu Wei, Jun Ma, Maarten de Rijke, Leveraging contextual sentence relations for extractive summarization using a neural attention model, in: Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2017, pp. 95–104.

[38] Changjian Fang, Dejun Mu, Zhenghong Deng, Zhiang Wu, Word-sentence co-ranking for automatic extractive text summarization, Expert Syst. Appl. 72 (2017) 189–195.

[39] Masaru Isonuma, Toru Fujino, Junichiro Mori, Yutaka Matsuo, Ichiro Sakata, Extractive summarization using multi-task learning with document classification, in: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, 2017, pp. 2101–2110.

[40] Shashi Narayan, Nikos Papasarantopoulos, Shay B. Cohen, Mirella Lapata, Neural extractive summarization with side information, 2017, arXiv preprint arXiv:1704.04530.

[41] Yong Zhang, Meng Joo Er, Rui Zhao, Mahardhika Pratama, Multiview convolutional neural networks for multidocument extractive summarization, IEEE Trans. Cybern. 47 (10) (2016) 3230–3242.

[42] Ramesh Nallapati, Bowen Zhou, Mingbo Ma, Classify or select: Neural architectures for extractive document summarization, 2016, arXiv preprint arXiv:1611.04244.

[43] R. Vale, R. Lins, R. Ferreira, Assessing sentence simplification methods applied to text summarization, in: 2018 7th Brazilian Conference on Intelligent Systems (BRACIS), 2018, 49–54.

[44] Rafaella Vale, Rafael Dueire Lins, Rafael Ferreira, An assessment of sentence simplification methods in extractive text summarization, in: Proceedings of the ACM Symposium on Document Engineering 2020, 2020, pp. 1–9.

[45] Alexander M Rush, Sumit Chopra, Jason Weston, A neural attention model for abstractive sentence summarization, 2015, arXiv preprint arXiv:1509.00685.

[46] Sumit Chopra, Michael Auli, Alexander M Rush, Abstractive sentence summarization with attentive recurrent neural networks, in: Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2016, pp. 93–98.

[47] Ramesh Nallapati, Bowen Zhou, Caglar Gulcehre, Bing Xiang, et al., Abstractive text summarization using sequence-to-sequence rnns and beyond, 2016, arXiv preprint arXiv:1602.06023.

[48] Qingyu Zhou, Nan Yang, Furu Wei, Ming Zhou, Selective encoding for abstractive sentence summarization, 2017, arXiv preprint arXiv:1704.07073.

[49] Yu Yan, Weizhen Qi, Yeyun Gong, Dayiheng Liu, Nan Duan, Jiusheng Chen, Ruofei Zhang, Ming Zhou, Prophetnet: Predicting future N-gram for sequence-to-sequence pre-training, 2020,

[50] Oriol Vinyals, Meire Fortunato, Navdeep Jaitly, Pointer networks, in: Advances in Neural Information Processing Systems, 2015, pp. 2692–2700.

[51] Abigail See, Peter J Liu, Christopher D Manning, Get to the point: Summarization with pointer-generator networks, 2017, arXiv preprint arXiv:1704.04368.

[52] Romain Paulus, Caiming Xiong, Richard Socher, A deep reinforced model for abstractive summarization, 2017, arXiv preprint arXiv:1705.04304.

[53] Asli Celikyilmaz, Antoine Bosselut, Xiaodong He, Yejin Choi, Deep communicating agents for abstractive summarization, 2018, arXiv preprint arXiv:1803.10357.

[54] Yen-Chun Chen, Mohit Bansal, Fast abstractive summarization with reinforce-selected sentence rewriting, 2018, arXiv preprint arXiv:1805.11080.

[55] Shuming Ma, Xu Sun, Junyang Lin, Xuancheng Ren, A hierarchical end-to-end model for jointly improving text summarization and sentiment classification, 2018, arXiv preprint arXiv:1805.01089.

[56] Tatiana Vodolazova, Elena Lloret, The impact of rule-based text generation on the quality of abstractive summaries, in: Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2019), 2019, pp. 1275–1284.

[57] Chin-Yew Lin, Rouge: A package for automatic evaluation of summaries ACL, in: Proceedings of Workshop on Text Summarization Branches Out Post Conference Workshop of ACL, 2004, pp. 2017–05.

[58] Claire Grover, Ben Hachey, Ian Hughson, The HOLJ Corpus. Supporting summarisation of legal texts, in: Proceedings of the 5th International Workshop on Linguistically Interpreted Corpora, 2004, pp. 47–54.

[59] Thomas K Landauer, Peter W. Foltz, Darrell Laham, An introduction to latent semantic analysis, Discourse Process. 25 (2–3) (1998) 259–284.

[60] K. Merchant, Y. Pande, NLP Based latent semantic analysis for legal text summarization, in: 2018 International Conference on Advances in Computing, Communications and Informatics, ICACCI, 2018, pp. 1803–1807, http://dx.doi.org/10.1109/ICACCI.2018.8554831.

[61] Deepali Jain, Malaya Dutta Borah, Anupam Biswas, Fine-tuning textrank for legal document summarization: A Bayesian optimization based approach, in: Forum for Information Retrieval Evaluation, 2020, pp. 41–48.

[62] Hongyan Jing, Sentence reduction for automatic text summarization, in: Sixth Applied Natural Language Processing Conference, 2000, pp. 310–315.

[63] Laura Manor, Junyi Jessy Li, Plain english summarization of contracts, 2019, arXiv preprint arXiv:1906.00424.

[64] Diego Feijo, Viviane Moreira, Summarizing legal rulings: Comparative experiments, in: Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2019), 2019, pp. 313–322.

[65] Sukriti Verma, Vagisha Nidhi, Extractive summarization using deep learning, 2017, arXiv preprint arXiv:1708.04439.

[66] Filippo Galgani, Paul Compton, Achim Hoffmann, Summarization based on bi-directional citation analysis, Inform. Process. Manage. 51 (1) (2015) 1–24.

[67] Filippo Galgani, Paul Compton, Achim Hoffmann, Citation based summarisation of legal texts, in: Pacific Rim International Conference on Artificial Intelligence, Springer, 2012, pp. 40–52.

[68] Atefeh Farzindar, Guy Lapalme, Letsum, an automatic legal text summarizing system, Legal Knowl. Inform. Syst., JURIX (2004) 11–18.

[69] Claire Grover, Ben Hachey, Ian Hughson, Chris Korycinski, Automatic summarisation of legal documents, in: Proceedings of the 9th International Conference on Artificial Intelligence and Law, 2003, pp. 243–251.

[70] M. Saravanan, Balaraman Ravindran, S. Raman, Improving legal document summarization using graphical models, Front. Artif. Intell. Appl. 152 (2006) 51.
[71] Ben Hachey, Claire Grover, A rhetorical status classifier for legal text summarisation, in: Text Summarization Branches Out, 2004, pp. 35–42.
[72] Simone Teufel, Marc Moens, Summarizing scientific articles: experiments with relevance and rhetorical status, Comput. Linguist. 28 (4) (2002) 409–445.
[73] Paheli Bhattacharya, Shounak Paul, Kripabandhu Ghosh, Saptarshi Ghosh, Adam Wyner, Identification of rhetorical roles of sentences in Indian legal judgments, 2019, arXiv preprint arXiv:1911.05405.
[74] Son Bao Pham, Achim Hoffmann, Incremental knowledge acquisition for building sophisticated information extraction systems with kaftie, in: International Conference on Practical Aspects of Knowledge Management, Springer, 2004, pp. 292–306.
[75] Filippo Galgani, Paul Compton, Achim Hoffmann, Combining different summarization techniques for legal text, in: Proceedings of the Workshop on Innovative Hybrid Approaches To the Processing of Textual Data, Association for Computational Linguistics, 2012, pp. 115–123.
[76] Xinyu Duan, Yating Zhang, Lin Yuan, Xin Zhou, Xiaozhong Liu, Tianyi Wang, Ruocheng Wang, Qiong Zhang, Changlong Sun, Fei Wu, Legal summarization for multi-role debate dialogue via controversy focus mining and multi-task learning, in: Proceedings of the 28th ACM International Conference on Information and Knowledge Management, in: CIKM '19, Association for Computing Machinery, 2019, pp. 1361–1370, https://doi.org/10.1145/3357384.3357940.
[77] Mi-Young Kim, Ying Xu, Randy Goebel, Summarization of legal texts with high cohesion and automatic compression rate, in: JSAI International Symposium on Artificial Intelligence, Springer, 2012, pp. 190–204.
[78] Ben Hachey, Claire Grover, Extractive summarisation of legal texts, Artif. Intell. Law 14 (4) (2006) 305–345.
[79] Ambedkar Kanapala, Srikanth Jannu, Rajendra Pamula, Summarization of legal judgments using gravitational search algorithm, Neural Comput. Appl. 31 (12) (2019) 8631–8639.
[80] Esmat Rashedi, Hossein Nezamabadi-Pour, Saeid Saryazdi, GSA: a gravitational search algorithm, Inform. Sci. 179 (13) (2009) 2232–2248.
[81] Tho Thi Ngoc Le, Minh Le Nguyen, Akira Shimazu, Unsupervised keyword extraction for Japanese legal documents, in: JURIX, 2013, pp. 97–106.
[82] Linwu Zhong, Ziyi Zhong, Zinian Zhao, Siyuan Wang, Kevin D. Ashley, Matthias Grabmair, Automatic summarization of legal decisions using iterative masking of predictive sentences, in: Proceedings of the Seventeenth International Conference on Artificial Intelligence and Law, 2019, pp. 163–172.
[83] Vu Tran, Minh Le Nguyen, Ken Satoh, Automatic catchphrase extraction from legal case documents via scoring using deep neural networks, 2018, arXiv preprint arXiv:1809.05219.
[84] Deepa Anand, Rupali Wagh, Effective deep learning approaches for summarization of legal texts, J. King Saud Univ.-Comput. Inform. Sci. (2019).
[85] Seth Polsley, Pooja Jhunjhunwala, Ruihong Huang, Casesummarizer: a system for automated summarization of legal texts, in: Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: System Demonstrations, 2016, pp. 258–262.
[86] Filippo Galgani, Achim Hoffmann, Lexa: Towards automatic legal citation classification, in: Australasian Joint Conference on Artificial Intelligence, Springer, 2010, pp. 445–454.
[87] Automatic text summarizer, 2020, http://autosummarizer.com/index.php (Online Accessed 03 March 2020).
[88] Umar Syed, Text summarizer, 2020, http://textsummarization.net/text-summarizer (Online Accessed 03 March 2020).
[89] Andreas Gohr, Open text summarizer, 2020, https://www.splitbrain.org/services/ots (Online Accessed 03 March 2020).
[90] SMMRY: online tool for summarization, 2020, https://smmry.com/ (Online Accessed 03 March 2020).
[91] Filippo Galgani, Paul Compton, Achim Hoffmann, HAUSS: Incrementally building a summarizer combining multiple techniques, Int. J. Hum.-Comput. Stud. 72 (7) (2014) 584–605.
[92] Anastassia Kornilova, Vlad Eidelman, Billsum: A corpus for automatic summarization of US legislation, 2019, arXiv preprint arXiv:1910.00523.
[93] Lawbriefs, 2020, https://lawbriefs.in/ (Online Accessed 20-August-2020).
[94] Cyber India blog, 2020, https://cyberblogindia.in/category/case-summary (Online Accessed 20 September 2020).
[95] Law times journal, 2020, https://lawtimesjournal.in/category/case-summary/ (Online Accessed 20 September 2020).
[96] News article 1, 2021, https://timesofindia.indiatimes.com/india/when-even-judges-cant-understand-judgments/articleshow/58690771.cms (Online Accessed 26 January 2021).
[97] News article 2, 2021, https://theprint.in/talk-point/supreme-courts-nearly-500-page-order-on-sec-377-are-indias-judges-most-erudite-or-too-verbose/113607/ (Online Accessed 26 January 2021).
[98] Emmanuel Chieze, Atefeh Farzindar, Guy Lapalme, An automatic system for summarization and information extraction of legal information, in: Semantic Processing of Legal Texts, Springer, 2010, pp. 216–234.
[99] Ambedkar Kanapala, Srikanth Jannu, Rajendra Pamula, Passage-based text summarization for legal information retrieval, Arab. J. Sci. Eng. 44 (11) (2019) 9159–9169.
[100] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, Jeffrey Dean, Distributed representations of words and phrases and their compositionality, 2013, arXiv preprint arXiv:1310.4546.
[101] Jeffrey Pennington, Richard Socher, Christopher D. Manning, Glove: Global vectors for word representation, in: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP, 2014, pp. 1532–1543.
[102] Ilias Chalkidis, Dimitrios Kampas, Deep learning in law: early adaptation and legal word embeddings trained on large corpora, Artif. Intell. Law 27 (2) (2019) 171–198.