A study on automatic detection method of the confidential word in the Japanese precedents

March 2021

By
Masakazu Kanazawa

Department of Innovation Systems Engineering
Graduate school of Engineering, Utsunomiya University,

Abstract

The globalization of the economy, international trade, and disputes internationally present new demands on judiciaries. At the same time, advances in information communications technology (ICT) offer opportunities for judicial policymakers to make justice more accessible, transparent, and effective. Jurisdictions in many countries have aimed to follow economic growth and political changes to allow anyone to easily access judiciary documents by incorporating ICT. Such justice systems empowered by ICT are called "cyber courts." In recent years, several studies on incorporating information communications technology (ICT) in courts have been reported. It is intended to allow anyone to easily participate in the trial. However, in the Japanese court system, certain personal names, company names, etc., are hidden in the judgment sentence due to privacy concerns. This anonymization work is done manually and requires a large amount of effort, resulting in less than 1% of cases being published.

To solve this problem, we introduced a technique to predict confidential words using a neural network automatically. In a previous study, we demonstrated that our method was effective in predicting the target words. However, the accuracy in predicting the confidential word was not good. Therefore, we attempted to improve the accuracy and found the part of speech (POS) of the

confidential words were a mostly proper noun. So, we proposed a new neural network model that combined POS tag extracting by MeCab, a Japanese morphological analyzer, and a kind of CRF (Conditional Random Field). We experimented with the new proposed model to detect confidential words and applied the method to anonymized precedents.

In this paper, we describe the mechanism of our proposed model and the prediction results using perplexity (PPL) which represented the number of prediction choices. Then, we evaluated how our proposed model was useful for the actual precedents by using recall, precision, and F1 score. As a first experimental result, we couldn't get satisfactory results predicting the confidential word. Then, we analyze the issue of the experiment, and doing the second experiment by improving the preprocessing algorithm, we were able to improve the value of CW_PPL, which means the average PPL value of predicting the confidential word in the test data, by 88% compared to a previous model.

Then, Appling the model to actual precedents (anonymized), about 60% of the cases showed a confidential word detection rate (recall score) of 70%–100%. Herein, we mention the result of applying this method to the plain precedents (not anonymized). However, the result was a recall rate of ≥70% was found in 33% of the total. We analyzed the result and found the reason for the low accuracy to detect confidential words. We think it may be possible to solve this problem to upgrade the POS function and performed a simulation. Then, we had the confidence to increase the accuracy to detect confidential words in a plain precedent. Furthermore, as a result of applying this model to actual cases, although the precision rate decreased, the recall rate was ≥70% in about 67% of the cases after anonymization. Issues encountered included the proposed model not predicting Greek letters and addresses, especially those containing a long number like "xx-xx-xx," as confidential words. By improving the preprocessing algorithm to recognize them as confidential words, the recall score was improved by 4%–49%. When we continued to experiment using more plain precedents after improving preprocessing, the recall rate was ≥70% in about 89% of the cases after anonymization. Although the precision is ≤30% and F1 is ≤50% in total, it is important to raise the recall score. It means the new proposed model has the potential for practical use.

So, the prediction tool that we will make using this model, should help reduce the current manual labor during anonymization, providing a step toward disclosure of all precedents.

Contents

Chapter 1

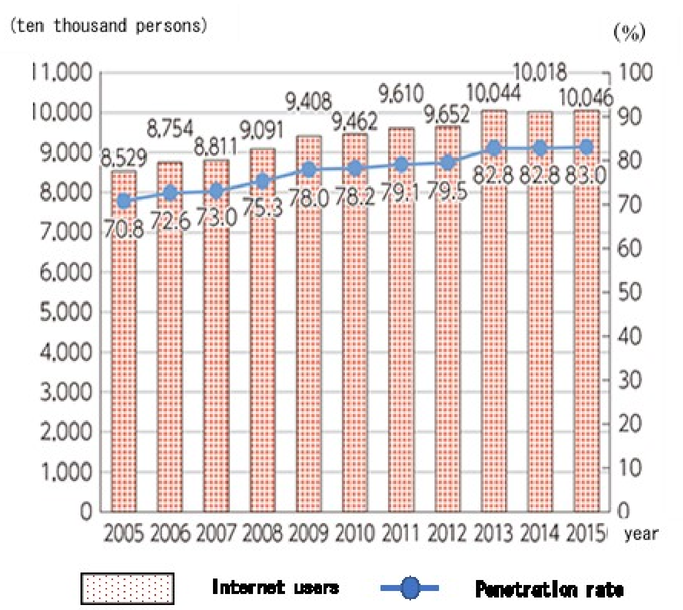
Introduction

In this chapter, I will describe the background of my study and its purpose.

Nowadays, Information and Communication Technology (ICT) are considered essential for our life. In the digital economy, the use of ICT has the potential to diversify business partners, expand trade areas, enable people to receive orders from distant locations, and make up for labor shortages using machines, etc.

To take advantage of these opportunities, developing ICT infrastructure, and making efforts to better utilize data are important. As ICT advances and expands, ICT is going beyond the online world and bringing huge transformations to real-world structures. The Number of internet users and the population penetration rate is shown in Fig.1.1 [1-1].

Consequently, the meaning of the digital economy has widened to include the new shape of the economy engendered by ICT, as well as the activities surrounding ICT in the overall economy.



(Source) "Communications Usage Trend Survey," MIC

Fig.1.1 Internet users and the population penetration rate

In this way, the evolution of ICT drives the evolution of the digital economy. Fig. 1.2 shows the ICT usage trend [1-1].

The business world is increasingly recognizing the Internet as a place where commercial dealings can take place, the legal system needs to be equipped to not only understand the Internet but also to use it as a tool to successfully exchange and use information. The Internet now provides a wide range of legal information, and one of the benefits of information being provided in this way is that it can be kept up-to-date as the law changes. Not only can the Internet assist in legal research but it can also assist in court processes in general, that is, in trial preparation and the courtroom throughout the hearing. Reasons that courts should embrace such technology lie in the constantly increasing caseloads, the complexity of cases and jurisdictions, resource constraints, the pressure to improve access to justice, expectations of performance improvements, and the pressure to improve efficiency and effectiveness in the court's administration and the delivery of justice. Therefore, jurisdictions of many countries aim to follow economic growth and political changes to allow anyone to easily access judiciary documents by introducing ICT (information and communications technology). Such a justice system empowered by ICT is called a “cyber court”.

The pioneer study of a cyber court system is Courtroom 21 [1-2]; which started in 1993 at the

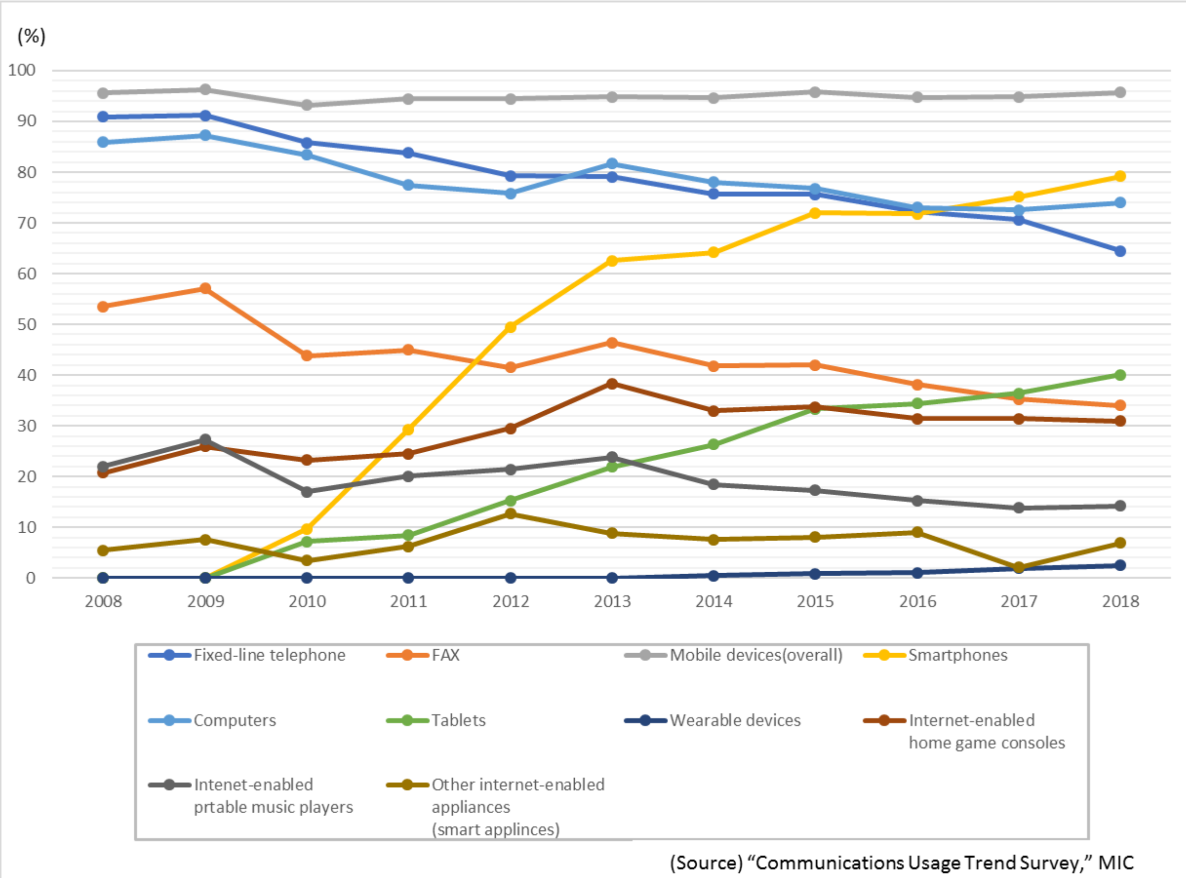


Fig.1.2 ICT user’s trend

College of William & Mary as a joint project between the university and the National Center for State Courts in the U.S.A. They implemented a cyber court using a teleconference system and performed multiple trials to determine the effectiveness of ICT in the court. Besides, the Singaporean Supreme Court introduced ICT in 1998 via measures such as a centralized display management system, a digital transcription system, e-signatures, electronic hearings, and an online justice system. All documents are displayed on PC screens in the court [1-3]. The High Court of Delhi and the District Court of Delhi have also introduced paperless e-courts. From the standpoint of technological research, many studies are focusing on models combining law and the knowledge structures of law [1-4], [1-5]. Even China began to stream some trials in more traditional courtrooms online in 2016 in an apparent effort to boost the transparency of their legal system [1-6]. Accordingly, we decided that it is very important to develop a prototype of a cyber court to evaluate the effectiveness of such an idea.

In Japan, the prototype for the first civil trial was developed at the Toin University of Yokohama in 2004 [1-7, 1-8], and its effectiveness, particularly its usefulness in the Saibanin system (the Japanese jury system), was proven [1-9]. An experiment with a remote trial was also conducted [1-10]. The Investments for the Future Strategy 2017 by the Japanese Cabinet Office includes ICT conversion for trials to accelerate them and improving their efficiency [1-11]. And also, reported the Future investment strategy 2018 about the IT conversion such as court procedure as described in Fig.1.3 [2-2].

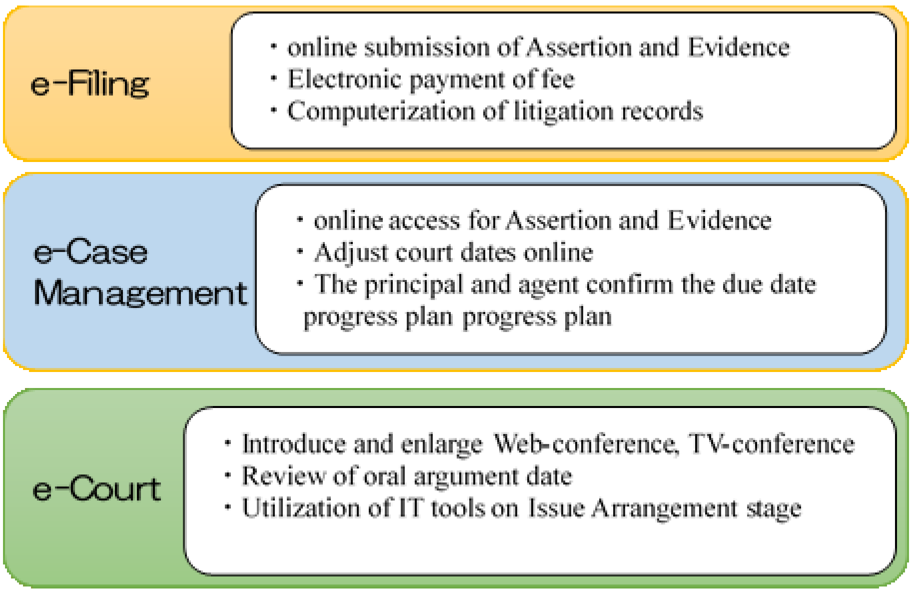


Fig.1.3 Main contents of converting court procedure to ICT

From the viewpoint of using big data, this field is expected to develop further. The disclosure of judicial precedents is indispensable for using big data in AI and the field of law; however, most precedents are not publicly open on Japanese court webpages. This is because Japanese privacy stipulates that an individual not be specified. Online privacy in Japan is primarily governed by general law, the Act on the Protection of Personal Information (APPI), rather than a specialized law for online privacy. The APPI applies to all business operators that possess individuals' personal information. Japan has other personal information protection laws that apply to government and public organizations [1-12].

Confidential words (e.g., personal, corporate, and place names) in opened precedents are replaced by other meaningless words, such as a single uppercase letter.

This operation takes time and effort because it is done manually. Besides, because trials involving people from various countries have been increasing due to globalization in recent years, creating a proper noun dictionary is difficult.

To solve this problem, some method has been considered. One method is to extract the confidential word using the Named Entity Recognition (NER). The NER extraction is executed roughly by two methods, one is rule-based by pattern matching, and the other is statistical-based by machine learning. Pattern matching takes very a high cost because of making the pattern of the named entity dictionary or updating them manually. Various kind of methods of machine learning is studied to solve the problem. The method used in machine learning can learn the pattern of a named entity by preparing the corpus. There are HMM (Hidden Markov Model), and CRF (Conditional random fields) for a machine language. CRF has mostly successful for NER. Nevertheless, the problem of machine learning is that the cost of manually making a corpus is high [1-13]

However, neural networks have advanced greatly in the field of natural language processing in recent years. Studies including deriving a vector considering the meaning of words and predicting words are actively ongoing [14]. Therefore, we proposed two neural network models: a bi-directional

LSTM (Long Short-term Memory) LR (left-right) model and a Sum-LSTM based on the CBOW (continuous bag-of-words) model. We experimented with these models to ascertain their effectiveness for predicting confidential words.

As a result, we were able to predict confidential words but did not obtain good accuracy. We review our experimental results and then we experiment with the proposed neural network by changing the window size and choosing proper parameter values. Then, we enhance the preprocessing of the input datasets. This operation is important for the neural network learning process. We evaluate the results of the experiment and propose a new model combined with a Part of Speech (POS) tag to form a more powerful method to improve accuracy. I experimented with the Japanese precedents using the new model. As a result, I could get better accuracy compared to the previous model. We got the CW_PPL score was 88% improved in accuracy and 20 % improvement for detecting the target word (PPL) compared with the previous model. I would explain the results of our study including the predicting results for the confidential words in Japanese actual precedents. As the result, we confirmed our proposed model (Bi-directional LSTM-LR combined CRF) had high accuracy for predicting the confidential words. As we got an excellent predicting ability with our proposed model, we need to confirm if it is practical or not in the next step.

By applying this method to actual precedents after anonymization, about 60% of the cases demonstrated a recall score (the ratio of correctly predicted positive observations to total tagged observations) of confidential words in the range of 70%–100%.

However, issues in some cases resulted in recall scores that decreased. In particular, some person names and long addresses weren't recognized as confidential words. The problems of some unrecognized addresses were solved by improving the preprocessing algorithm. Fig.1.4 shows the outline of this paper.

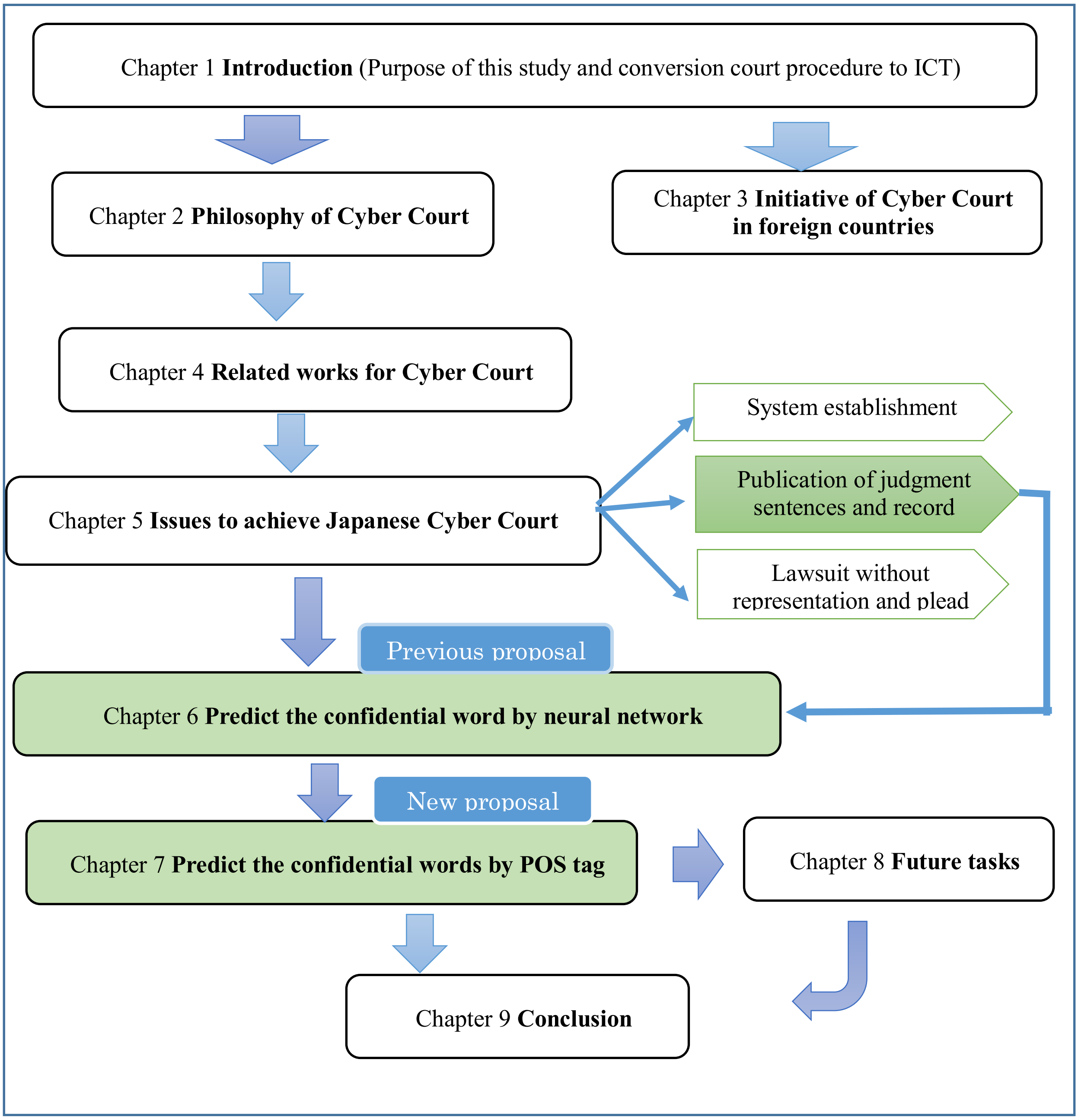


Fig.1.4 Outline of this paper

In chapter 1, the purpose of this study, and the outline of this paper is described. In chapter 2, the fundamental philosophy of the Cyber Court is described, and how it influences civil life is explained. Besides, the concrete contents of the Cyber Court are also described. In chapter 3, Initiative Cyber Court in the main foreign countries is explained. In chapter 4, I investigated the status and issues of using ICT in the Japanese court system. One of the issues is privacy. In chapter5, the issues to achieve the Japanese Cyber Court are clarified. In chapter 6, we describe the predicting method of confidential words in Japanese precedents including the related works and the experimental results. Then, in chapter 7, we propose the new prediction model using the neural network combined POS tag to improve the accuracy of predicting the confidential word. And I describe the results of the experiment. Then, I clarify the issues of the experimental results and show the countermeasures for them. After that, improving the preprocessing algorithm as a countermeasure I notice the experimental results that the proposed model has shown the possibility of practical use. In chapter 8, I explain the future work. In chapter 9, I conclude my study.

Chapter 2

Philosophy of Cyber Court

2.1 Basic principle

The right to a trial in law for all humans must be fair and open and formally accessible to justice. But it isn't impartial to access justice because of the difference of knowledge about the law among humans and the distance to the court. The Cyber Court will solve this problem. The goal and legitimacy of civilization lawsuits must be based on the realization and convenience of citizens' right to a trial, and the quality of civil proceedings. Besides, the "IT conversion of civil litigation" goes beyond such convenience and contributes to enhancing and speeding up the examination of civil litigation. Civil litigation must be easy to understand and use for the people and is reliable. [2-1]

Cyber Court, the so-called "e-Support System" considers itself to be at the heart of the upcoming comprehensive justice system reform [2-2].

Cyber Court refers to the entire system that electronically processes and manages all litigation cases in civil litigation. In other words, starting from a civil case, which is a source of information about a civil trial, a complaint, an answer sheet, and a preparation Litigation materials and evidence materials contained in documents and evidence, information on the progress of the litigation, schedules and other trial records of the judge/court clerk in charge, all information on judgments and execution/preservation, and various other information a system for electronically processing and managing information that facilitates sharing and utilization of information (meta-information necessary for information processing) for each case [2-3]. Fig.2.1 shows a conceptual diagram of the Cyber Court [2-4].

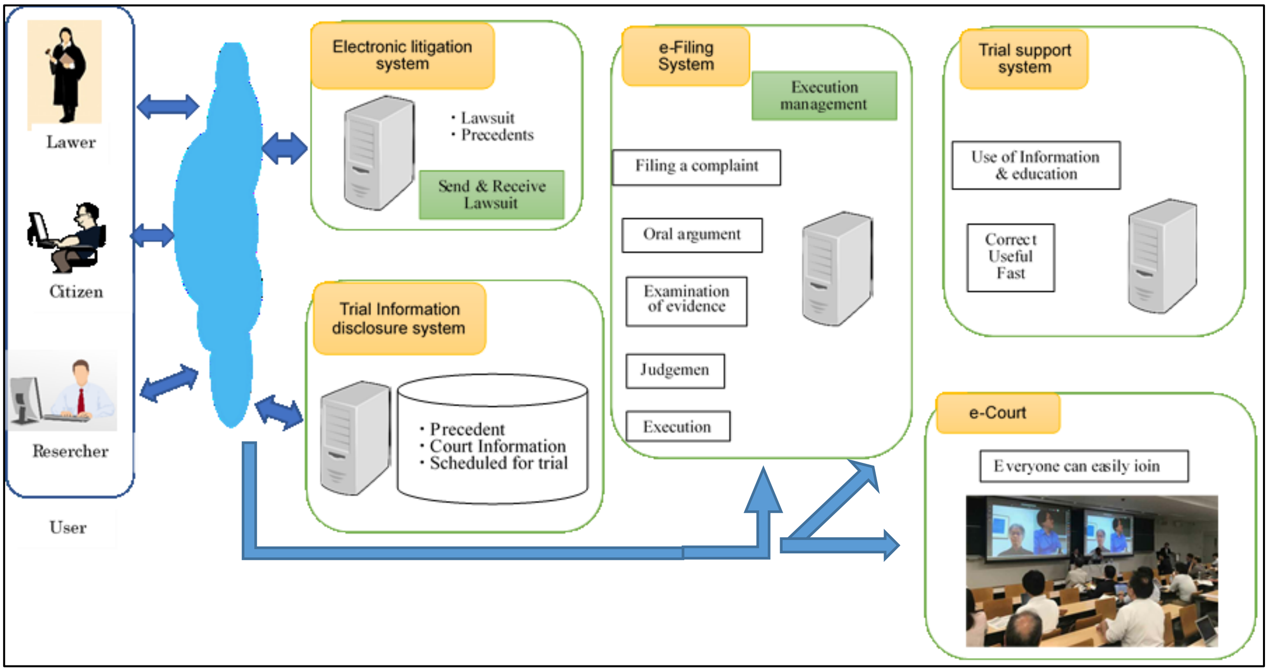


Fig.2.1 Conceptual diagram of Cyber Court

2.2 Main contents of Cyber Court

2.2.1 e-Filing

"e-Filing" is that instead of the current handling of bringing paper documents to the court and mailing them to courts, we will shift as much as possible to online submission by electronic information, which is available 24 hours a day, 365 days a year, and will unify it.

2.2.2 e-Case Management

“e-Case Management” means that both the litigant and the litigation agent can easily and at any time access the electronic information such as a complaint, answer, or other preparatory documents or evidence, to the case records and case information managed by the court. It will be possible to access online, and a mechanism that can check the progress of the due date, etc. will be established. This will increase the transparency of the court proceedings and can be expected to free the parties themselves and their agents from the burden of bringing and keeping paper-based litigation records on their own.

2.2.3 e-Court

The e-court is the entire process of civil litigation in to reduce the time and economic burden of appearing in the courts of the parties, etc., and to enhance the quality of trials with a due date. Significantly increase the use of video and web conferencing by one or both parties.

Chapter 3

An initiative of Cyber Court in foreign countries

3.1 Introduction

In countries such as South Korea, Singapore, and the United States, the use of ICT in courts has been actively introduced or is under consideration. However, in Japan, introduce of ICT for legal field is too late rather than these countries. In this section, I will describe the status of Cyber Court in foreign countries. [3-1]

3.2 USA

Trial Information Access System PACER (Public Access to Court Electronic Records) launched in 1996. Using this system everyone can view and download the all case records pending federal court. The CM / ECF (The Case Management /Electronic Case Filing) is introduced in the late 1990s. The CM / ECF is electronic filing system for filing cases and providing documents to federal courts. As a result of adopting the electronic filing system, lawyers can submit papers on the web, and judge and lawyer can see the documents on the web from anywhere [3-2]. About the use of ICT in court proceedings, it is possible to do the remote appearance by using a system or a telephone conference system to conduct video conferencing in the process of organizing issues and oral arguments. Courtroom 21 Project at William Mary University is famous case study.

3.3 Singapore

In Singapore, a civil action has been filed since March 2000. All lawsuits required electronic filing on paper. No complaints will be accepted. This system is called EFS (Electronic Filing System).

Singapore's EFS is governed by the laws of both courts and plaintiffs / defendants. It is intended for use between these offices. Current court proceeding, Electronic Litigation System (eLit) is used since 2013. “e-Filing”, “e-Case Management” and “e-Court” are available using this system [3-2].

In future, Singapore will plan to introduce the Online Dispute Resolution (ODR) system. It has “Outcome Simulator” that is a simulator that analyzes the data of past court cases and predicts the amount of claims so that the parties can predict the amount of damages to be paid before trial or ODR.

3.4 Korea

An electronic trial system has been implemented in the Patent Court, which seeks a cancellation action against a decision of the Intellectual Property Office in 2010. From 2011, an electronic trial procedure was applied to a general civil trial. All legal records, including evidence, are kept electronically. No paper lawsuit record is made. The purpose of electronic trials is to provide courts benefits that the efficiency of paperwork is improved and the convenience of judicial services is improved through the digitization of court procedures.

For e-Case Management, everyone can view information on related cases of the person in charge in the form of a portal site. Specifically, "List of raised cases", "Progress of each case", "List of documents delivered to the person", etc. Regarding oral arguments conducted by the court, procedures by a web conference or video conference (remote video trial) have not yet been implemented. However, some courts are conducting trial preparatory proceedings on a remote video basis to verify problems [3-4].

3.5 China

From July 1, 2016, all public trials in the Supreme Court of the People's Court (equivalent to the Supreme Court) will be broadcast live on the Internet in principle.

A cyber court was established in Hangzhou in August 2017. Although this applies only to some cases such as e-commerce disputes, it is possible to use a web conference for hearings and to do everything from complaints to judgments online [3-5].

3.6 France

In response to a people's request for free release of legal information via the internet from the guarantee of the right of people to know, the website "Legifrance" was opened in 19997. People can free access to an online database on the website "Legifrance".

Réseau privé virtuel justice (RPVJ), that is an intranet system managed by the Ministry of Justice, has been constructed and used as the basis for civil, criminal and administrative court proceedings. And Lawyers do not have direct access to the RPVJ, but instead have access to a network of Réseau privé virtuel avocats (RPVA) operated by Conseil National des Barreaux, a national council for bar associations. The RPVA is interconnected with the RPVJ on the court side, enabling mutual electronic communication [3-6].

3.7 Germany

The use of ICT in trials in Germany has been delayed in the EU. The law on the Use of Electronic Information Forms in Justice has enacted in 2005. Comprehensive electronic information exchange (e-filing) has been enabled by the law. Only the electronic documents with a qualified electronic signature can be recognized as documents in court. Electronic information exchange (e-filing) in all courts has been enabled in 2018.

A portal service that allows people to view the trial records online will be created at the state level in the future [3-3].

.

3.8 Spain

Civil procedure law has been amended to obligate the voice and video recording in 2000. By revision of the Justice Ministry legislative process in 2009, court recordings can apply to all jurisdictions and record data is approved by the secretary and stored with electronic signature. Remote access to courts saves time and money and improved court security by eliminating defendants' movements and allowing them to participate remotely [3-3].

3.9 UK

Not all civil cases have achieved IT conversion in trial. However, online claims has been used since November 2015 in Rolls Building in London. By using CE-File (e-filing), it is possible to submit various documents, including filing a complaint and filing a counterclaim. CE-File is used for filing an appeal, counterclaim, petition, etc. It is a system that can be submitted 24 hours a day, 365 days a year. In the UK, litigation is also permitted, so the parties themselves can use this system. According to the filing procedures, all records will be submitted using the CE-File system and will be kept in court. Regarding the inspection of the record, the party's representative viewed it at the court terminal. It appears that electronic copies can be obtained, and third parties can view and request electronic records within Rolls Building. [3-7]

“Legal Aid, Sentencing & Punishment of Off enders Act 2012” was newly established in 2012. The provision of legal assistance services was a state responsibility, but provided that the service would be reduced if provided by telephone or other electronic devices. Alternative Business Structure

(ABS) was introduced under the Legal Services Act 2007 as a new service model due to deregulation of legal services. This means that even non-legal workers can own and manage law firms. This is a system that allows participation in the market for both users of legal services and providers of legal services1. Co-op group has advanced into legal services using ABS in 2011. Since September 2012, the company has provided civil affairs services centering on housework, labor, inheritance, etc., utilizing the nationwide network of Co-op.

Chapter 4

Related works for Cyber Court

4.1 Introduction

As above described, electronic lawsuits in civil cases have been filed using IT in Korea since May 2011, and it is now being used in most cases. On the other hand, in Japan, court procedures were set up in the Cabinet Secretariat in the wake of the fact that the introduction of IT in court procedures, etc. was indicated as an issue to be considered in the "Future Investment Strategy 2017" decided by the Cabinet in June 2017. In March 2018, the IT Study Committee published a compilation document, indicating the direction to pursue full-scale IT such as court proceedings.

However, according to the report, the electronic submission of litigation documents and the electronic access to litigation records, which have been realized in Korea, will be realized in around FY2022, while it is said that it is desirable to consider the schedule for realization, but no specific realization schedule is given. Further, World Bank "Doing Business" (Note: World Bank announces annually selected and prioritized 10 areas related to business activity regulation in 190 countries) in the 2017 edition, strict evaluations were given to Japan regarding items related to "automated court procedures (IT)". The ease of doing business ranking in 2019 is shown in Table 4.1.

Table 4.1 Ease of doing business ranking

Rank	Economy	Rank	Economy
1	New Zealand	21	Iceland
2	Singapore	22	Canada
3	Denmark	23	Ireland
4	Hong Kong SAR, China	24	Germany
5	Korea, Rep.	25	Azerbaijan
6	Georgia	26	Austria
7	Norway	27	Thailand
8	United States	28	Kazakhstan
9	United Kingdom	29	Rwanda
10	Macedonia, FYR	30	Spain
11	United Arab Emirates	31	Russian Federation
12	Sweden	32	France
13	Taiwan, China	33	Poland
14	Lithuania	34	Portugal
15	Malaysia	35	Czech Republic
16	Estonia	36	Netherlands
17	Finland	37	Belarus
18	Australia	38	Switzerland
19	Latvia	**39**	**Japan**
20	Mauritius	40	Slovenia

Source: Doing Business 2019

Table 4.1 shows the ease of business environment in Japan is inferior to the main developed countries.

From the viewpoint of Japan's business environment and international competitiveness, there has been increasing demand that it is necessary to further promote the use of IT in court procedures from the perspective of users. Under such circumstances, the Government's “Future Investment Strategy 2017” (decided by the Cabinet on June 9, 2017) stated, “To achieve a prompt and efficient trial, from a comprehensive point of view, including procedural security and information security, with the cooperation of related organizations, etc., promptly examine measures to promote the use of IT in procedures related to courts from the user's perspective and reach a conclusion within this fiscal year. Following this guideline, the “IT Study Group on Trial Procedures, etc.” (The Study Group) was established in October 2017.

4.2 Needs and basic direction of court procedures using ICT

4.2.1 Needs for ICT in court procedures

Business people said that video and web conferences by widespread use in Japan, the time and economic burden of appearing in a distant court is expected to be reduced and court procedures will be quicker and more efficient. And the digitization of litigation records, which are premised on paper media, will make it easier to understand the case and reduce the cost of keeping records.

Many lawyers use IT tools, including interacting with clients. The use of IT to reduce the burden of appearance duties is reduced by the fact that the agent has an environment that can respond to the use of IT in court procedures and that IT equipment is used effectively. Then, they can make a well-informed hearing and utilizing claims/evidence as electronic information for ease of organizing/searching, etc., they can be contributing to speeding up and enhancing court proceedings.

From the standpoint of not only lawyers and other legal experts, but also businesses and consumers, there is great expectation for the promotion of IT in court procedures, etc. People have very strong needs for the conversion to ICT in court procedure. Also, with the rapid development of information and communication technology, the trend of the society where the improvement of the environment and the improvement of convenience are progressing further, the realization of full-scale IT is strongly expected in Japanese court procedures. It can be said that there is no waiting to take immediate action toward that.

4.2.2 Basic direction of ICT conversion

The use of IT in court procedures, etc., must properly reflect the development and penetration of information and communication technology in modern society, to realize appropriate, prompt, and easy-to-use trials. It should also be promoted from a comprehensive perspective such as strengthening the international competitiveness of dispute resolution infrastructure, rationalizing administrative burdens related to trials, and cost-effectiveness. Beyond the framework of the current law, from the filing and filing of a complaint to the subsequent procedure, it is necessary to consider a fundamental response to IT adoption that does not consider the existence of paper media.

As the basic direction of IT adoption in court procedures, it is important aiming for full-scale IT in court procedures, standing and premised on the complete digitization of litigation records. Considering the current situation in Japan, where the proportion of litigation in litigation cases is quite high, it is also important to further improve the people's access to justice with the adoption of IT, from the viewpoint of substantially guaranteeing the right of citizens to be tried.

4.3 Realization of “3-e”

When considering the use of IT in court procedures, etc., it is necessary to consider the needs of users and the situation in other countries, etc. That is, we must aim for the realization of 3-e (①e-Filing, ②e-Court,③e-case Management). The current civil lawsuit procedure is shown in Fig.4.1.

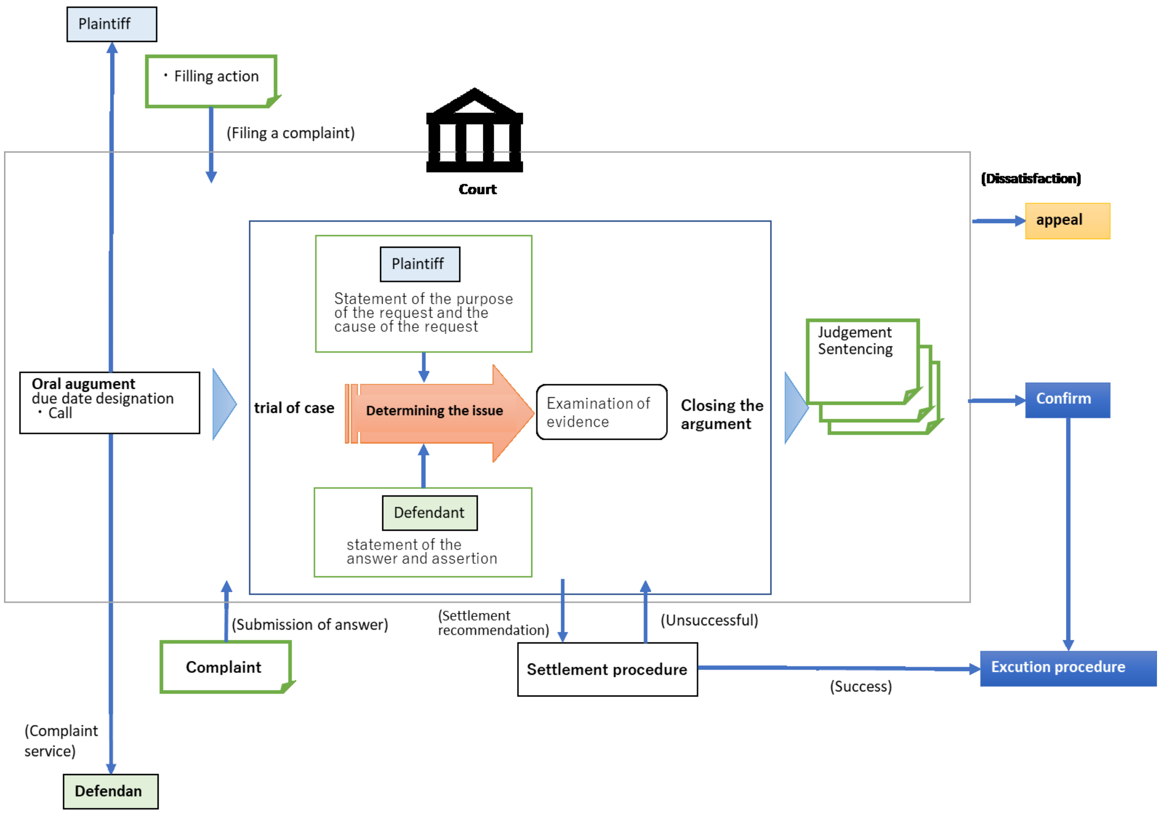


Fig.4.1 Current civil lawsuit procedure flow

4.3.1 e-Filing

From the user's point of view, electronic filing (e-Filing) is available online 24 hours a day, 365 days a year, instead of the current handling of bringing paper documents to the court and mailing them to courts. It is desirable to shift to online submission by electronic information as much as possible and unify it (do not use paper media for litigation records). Instead of the current treatment of filing paper complaints with the court, it will move to file online complaints (including the digitization of paper-based complaints). It is necessary to reconsider the current handling of the court, which in principle will deliver documents by mail, from the viewpoint of promoting the digitization of legal records, minimizing the coexistence of electronic information and paper media, and promoting online use.

Regarding the submission of a response, etc. from the defendant, the subsequent submission of written documents by both parties, and the exchange between the parties, it is necessary to consider measures to make online promptly and efficiently as well.

4.3.2 e-Case Management

Regarding case records and case information managed by the court, both the litigant and the litigant will be able to access electronic information such as complaints, defenses, and other preparatory documents and evidence online at any time and easily. It is desirable to establish a mechanism that can also check the progress of the due date by using IT.

Plaintiffs filed a complaint filed online that was accepted by the court need a mechanism that can reliably and easily confirm.

It is also necessary to use IT tools to examine prompt and efficient measures for examining court complaints and exchanging cases where amendments are required. It is useful for both the plaintiff and the defendant and the court to coordinate the progress schedule online including the due date etc.

4.3.2 e-Court

①Summary

From the user's point of view, to reduce the time and economic burden of appearing in the courts of the parties, etc. It is desirable to greatly expand the use of one or both video conferences and web conferences.

②First due date

Web conferencing by one or both parties (for example, at the nearest court or lawyer's office) to conduct a substantive hearing, and to effectively utilize the litigation record that has become electronic information. It is necessary to carry out a prerequisite hearing. Also, if there is no dispute in the contents of the claim or the defendant does not respond, the effective use of the web conference, etc. will promptly lead to the settlement procedure and the judgment procedure without the burden of the parties appearing.

③Issue of the resolution procedure

It is necessary to consider ways to use IT tools such as web conferences so that parties etc. can be involved in arranging issues without actually appearing in court.

Besides, from the viewpoint of high convenience, online in a place other than the court (for example, a lawyer's office, a meeting room of a company, or an appropriate space such as a public institution with a window for citizens). It is necessary to take new measures to make it possible to meet the deadline.

④Examination date

In the case of reviewing the current handling of videoconferencing and judging that it is necessary and appropriate by the court, one or both parties, witnesses, and other related parties will not go to the court and will be able to access the nearest lawyer's office or corporate meeting room, etc. It is necessary to conduct personal and witness interrogations at web conferences.

⑤Judgment sentence

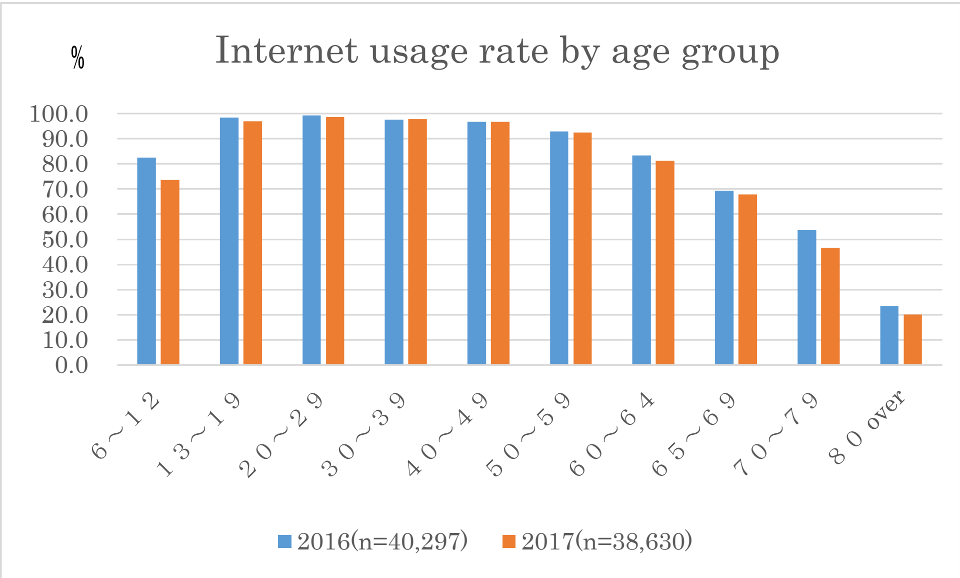
From the user's point of view, it is considered necessary to consider a framework for giving judgmental information, which is currently in paper form, the originality of the judgment information, which is electronic, following the conversion of litigation records to electronic information. Regarding the due date of the judgment, consider ways to respond to the needs of the parties, while paying attention to the public rules of the trial, etc.

Chapter 5

Issues to achieve of Japanese Cyber Court

5.1 Introduction

The fact that courts are accessible to citizens is an essential requirement to substantiate the Constitution's guaranteed right to trial (Article 32 of the Constitution). Even if the use of IT is a trend in society, too much emphasis is placed on pursuing the convenience of the majority, and measures to leave those who are not proficient in IT are necessary from the viewpoint of guaranteeing the right to trial. According to the Internet usage by age group (individual) in the Ministry of Internal Affairs and Communications' 2017 Telecommunications Usage Trends Report (Households), IT usage tends to decrease as the elderly and the financially disadvantaged. Fig 5.1 shows an internet usage rate by age group. Those who are not familiar with IT may be deprived of the opportunity to use civil trials. In civil court proceedings, it is necessary to ensure that the parties have the opportunity to assert and prove. However, the introduction of IT in civil courts is inconvenient for elderly people who are not proficient in IT, persons with disabilities, and those who conduct litigation. It is necessary to create a system that poses a danger and does not cause such a situation.



(Source) Communication usage survey in 2018 by MIC

Fig.5.1 Internet usage rate by age

5.2 System establishment (Standardization, generalization, and globalization of systems)

For establishing an e-filling system, judicial materials, judgment sentences, voice data, etc., are must standardize these data totally to use easily and develop them. One resolution is using the XML language. Since XML can define the description method of a document, the content of the document and the document information can be separated and integrated into the data. It is also suitable for exchanging information via the Internet, and standardization work is ongoing worldwide in many fields. In adopting the electronic court procedure system, it is necessary to consider the globalization of Japanese court procedures, such as the execution of foreign court procedures, English court procedures, and the publication of Japanese court databases overseas. As a means of standardization and globalization, the adoption of XML adopted by the Japanese Patent Office is considered essential.

5.3 Security measures

In promoting IT in court procedures, it is necessary to take sufficient measures from the viewpoint of information security. To prevent spoofing, falsification of documents, and denial of transmission (denying a document once submitted is not your submission), the adoption of electronic authentication is important. [5-1]

The required information security level and information security measures (identification, prevention of falsification/leakage, etc.) differ depending on the procedural stages of litigation and the content and nature of information such as litigation records, etc.

5.4 Publication of judgment sentences and record

Regarding the inspection of the case record by a person other than the parties, the current law allows any person to request the court clerk to inspect the case record (Article 91 (1) of the Act).

If the disclosure of oral arguments is prohibited, it can be requested only by a third party who has clarified the interest (paragraph 2 of the same article). In cases where full-scale digitization of litigation records is realized, Regarding the inspection of the case record by the party, the same shall be the case with the parties, except that prohibition of oral argument is prohibited, that the parties shall be able to access the case management system freely and that the inspection of the case record shall be permitted.

However, litigation records contain a lot of private information on related parties. For this reason, there is a problem in making this available on the Internet and making it freely available for viewing, and there is not much resistance even in Japan's public awareness.

Besides, even if the rule that only the lawsuit record can be viewed and that copying (downloading) is not allowed is adopted, if screenshots are taken, the data becomes easily distributable and the privacy of the parties may be harmed.

Therefore, in the current Japanese precedents, confidential words are replaced with another meaningless character by humans to protect privacy. It takes more effort and time.

For this reason, the disclosure rate of Supreme Court judgments is extremely low at around 0.9% [5-2].

So, we would study a method to automatically detect the confidential words that should be made invisible in legal documents.

The detail of our study is described next sections.

Chapter 6

Predict the confidential word by neural network

6.1 Introduction

Cognitive Info-communications (CogInfoCom) [6-1] [6-2] describes communications, especially the combination of informatics and communications. Future infocommunication is expected to be more intelligent and would even have the ability to support life. Fig. 6.1 shows the idea of CogInfoCom. Privacy is one of the most critical concerns in infocommunication. Encryption is a well-recognized technology used for ensuring privacy; however, encryption does not effectively hide personal information completely. One technique to protect privacy is to find confidential words in a file or a website and convert them into meaningless words. It will be good if a network becomes intelligent and automatically changes private words into meaningless words. We think that this is one benefit of introducing cognitive infocommunication into our life.

Based on a Japanese judicial precedent dataset, I discuss a recognition technique of confidential words using neural networks. The disclosure of judicial precedents is indispensable for using big data in the artificial intelligence (AI) and law field, but most precedents are not available publicly on the Web pages of Japanese courts.

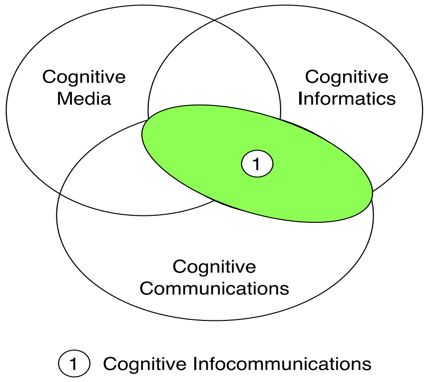


Fig. 6. 1. Infocommunication model

. This is because specifying any individual's name or other personal details violates the Japanese understanding of privacy. Confidential words, such as personal names, corporate names, and place names, in the precedents available for public viewing, are converted into other words to protect privacy. In Japan, this procedure takes time and effort because it is done manually. Also, globalization has led to the participation of people from various countries in these trials; therefore, a dictionary of proper nouns would take additional time to create.

Neural networks are also being increasingly used in natural language processing in recent years. There is ongoing research to predict words and to derive a vector based on the meanings of words [6-3].

Therefore, in this paper, I discuss ways of applying a neural network to the task of detecting confidential words.

6.2 Converting words

6.2.1 Confidential words in Japanese judicial precedents dataset

Some judicial precedents datasets are available for free on the websites of the Japanese courts [6-4]. Confidential words in these precedents that are available for public viewing on the website are converted into uppercase letters. (In paid magazines and websites, Japanese letters are sometimes used.) Figure 6.3 shows an example of such changed words.

In the example shown in Fig. 6.3, the personal name "Kiryu" is substituted by the letter "A." If some more words need to be kept confidential, other letters of the alphabet are used (e.g., B, C, and D).

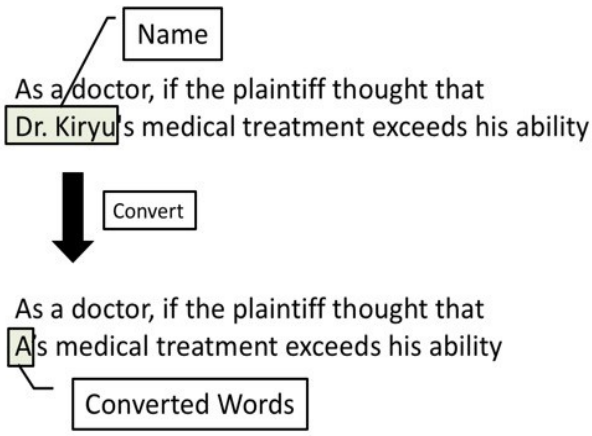


Fig.6.3 Example of converted words

In a current way, this action is doing manually. So, it takes many time and effort by a human.

6.3 Related works

To solve this problem, some method has been considered. One method is to extract the confidential word using the Named Entity Recognition (NER).

Conventional proper noun extraction is performed using a named entity dictionary. The named entity extraction API extracts a named entity such as a person's name, a place name, or a date representation from the Japanese character string sent in the request.

The named entity extraction is a widely used technique to get the target word in a sentence. Named entity recognition (NER) is probably the first step for information extraction to seeks to locate and classify named entities in text into predefined categories such as the names of persons, organizations, locations, expressions of times, quantities, monetary values, percentages, etc. NER is used in many fields in Natural Language Processing (NLP) [6-5] [6-6].

In general, named entity recognition (NER) is a technique for extracting personal names, place names, organization names, etc. from the text. It is applied to information retrieval, relation extraction, co-reference analysis, etc. The NER extraction is executed roughly by two methods, one is rule-based by pattern matching, and the other is statistical-based by machine learning.

6.3.1 A rule-based by pattern matching

A rule-based pattern matching is that humans create rules that determine how to determine which words correspond to which dictionaries depending on the context when the words appear. Each time a new word comes out, it must be created by humans. So, pattern matching takes a very high cost because of making the pattern of the named entity dictionary or updating them manually.

Then, we manually search for confidential words. These named entity words are useful for detecting confidential words. We provide an overview of this method in Fig. 6.2. This method eventually obtains confidential words manually. But the problem with this method is that we make the dictionary and always up to date the words in the dictionary because the new words appear day by day.

6.3.2 Statistical-based by machine learning

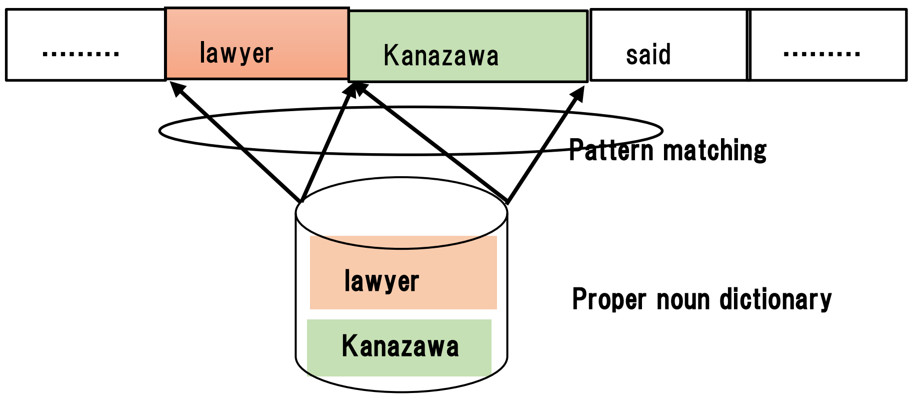


Fig.6.2 the existing method of extracting the proper noun

Various kinds of methods of machine learning are studied to solve the problem. The method used in machine learning can learn the pattern of a named entity by preparing the corpus. There are HMM (Hidden Markov Model), CRF (Conditional random fields), and SVM (Support Vector Machine) for a machine language.

(1)HMM (Hidden Markov Model)

HMM has a strong statistical foundation with efficient learning algorithms where learning can take

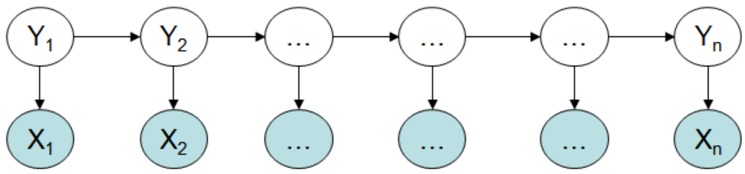


Fig.6.3 HMM model

place directly from raw sequence data. It allows consistent treatment of insertion and deletion penalties in the form of locally learnable methods and can handle inputs of variable length. They are the most flexible generalization of sequence profiles. It can also perform a wide variety of operations including multiple alignments, data mining and classification, structural analysis, and pattern discovery. It is also easy to combine into libraries.

Fig.6.3 shows an HMM model. Each state depends on the previous state. In other words (x_i,y_i) depends on (x_{i-1},y_{i-1})

$\boldsymbol{X}$ (X_1 ········· $X_{n:}$) **:** sequence (words)

$\boldsymbol{Y}$ (Y_1 ········· Y_n) **:** label sequence (Part of Speech).

P(x,y) : probability

$y^{predict}$: output the prediction parameter

.

$$P(x,y) = \prod_i P(x_i|y_i)P(y_i|y_{i-1}) \qquad (6\text{-}1)$$

$$y^{predict} = arg\max_y P(x,y) \qquad (6\text{-}2)$$

However, HMM is only dependent on every state and its corresponding observed object. The HMMs-based method is still limited because it is difficult to model arbitrary, dependent features of the input word sequence.

(2) CRF (Conditional random fields)

Conditional random fields (CRF) offer several advantages over hidden Markov models and stochastic grammars for such tasks, including the ability to relax strong independence assumptions made in those models. Conditional random fields also avoid a fundamental limitation of maximum entropy Markov models (MEMMs) and other discriminative Markov models based on directed graphical models, which can be biased towards states with few successor states [6-20].

CRF has mostly successful for NER. Nevertheless, the problem of machine learning is that the cost of manually making a corpus is high. CRF model is described in Fig.6.4

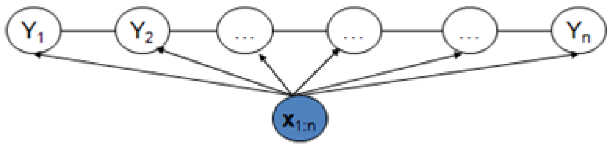


Fig.6.4 CRF model

X （X_1 ········· X_n ）**:** sequence (words)

Y （Y_1 ……… Y_n ）**:** label sequence (Part of Speech).

P(x,y) : probability

w : weight

ϕ : feature

$y^{predict}$: output the prediction parameter

$$P(x, y) = \frac{1}{Z_{x,w}} \exp(w \cdot \phi(x, y)) \qquad (6\text{-}3)$$

$$Z_{x,w} = \sum_y \exp(w \cdot \phi(x, y) \qquad (6\text{-}4)$$

$$y^{predict} = arg \max_y w \cdot \Phi(x, y) \qquad (6\text{-}5)$$

Since CRF does not have as strict independence assumptions as HMM does, it can accommodate any context information. Its feature design is flexible. However, CRF is highly computationally complex at the training stage of the algorithm. It makes it very difficult to re-train the model when newer data becomes available. An example of CRF is described in Fig.6.5.

CRF has proved to be quite successful for NER. Nevertheless, the problem of machine learning is that the cost of manually making a corpus is quite high [6-7].

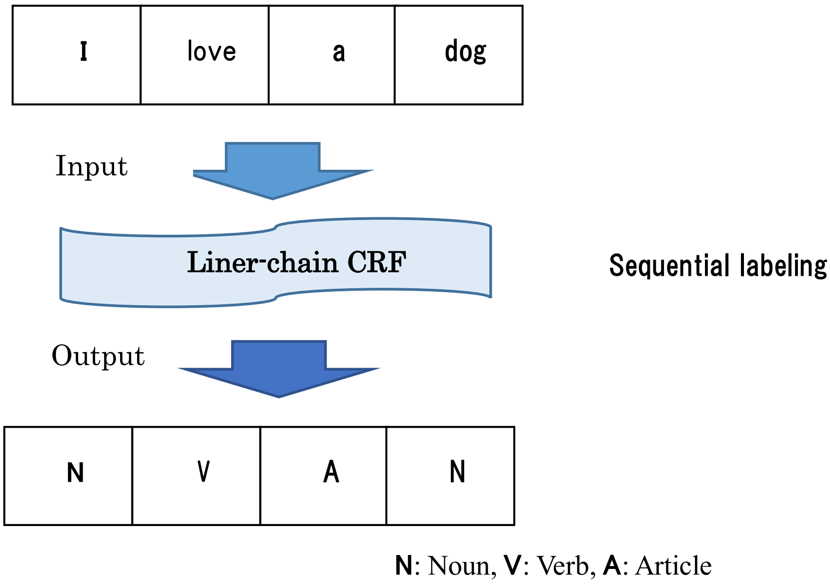


Fig.6.5 An example of CRF

(3) SVM (Support Vector Machine)

Support vector machines (SVMs) are a set of supervised learning methods used for classification, regression, and outliers detection. Support vector machine is highly preferred by many as it produces significant accuracy with less computation power. The generalization performance of SVM doesn't depend on the size of the dimension of feature spaces, such as named entity extraction task using lexical entries. SVM is still effective in cases where several dimensions are greater than the number of samples. However, the problem of SVM is that as the amount of learning data increases, the amount of calculation becomes enormous. Fig.6.6 shows the outline of SVM.

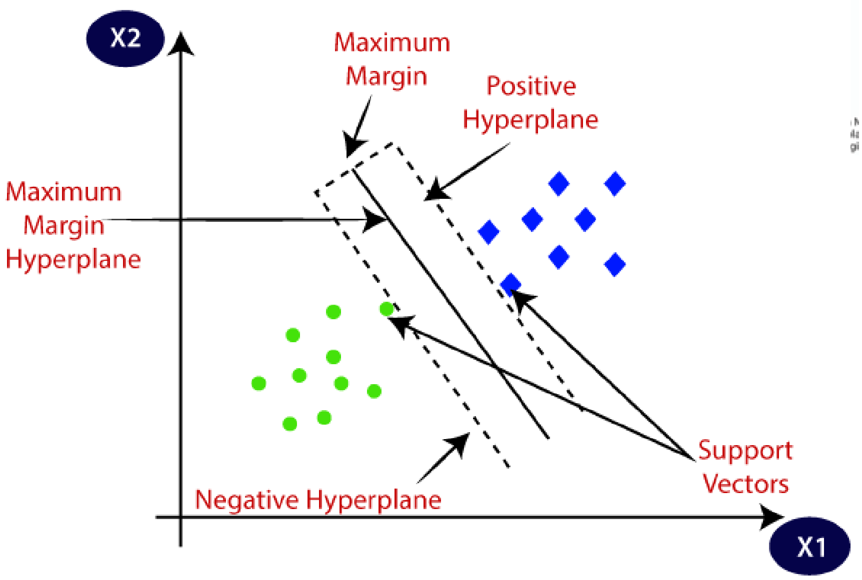


Fig.6.6 SVM model

6.3.3 Problem of detecting confidential words using a dictionary

Studies are also being conducted in other fields on various methods of hiding confidential proper nouns. The primary method is to create proper noun dictionaries, such as personal name dictionaries or place name dictionaries, and match them with the target documents.

The merit of this method is that the more the dictionary is enhanced, the more the accuracy improves. However, in Japanese, it is sometimes uncertain whether a word is a name or not unless it is read in the context of the entire sentence. Also, as globalization progresses, foreigners often join in trials, and it is difficult to create a dictionary that includes names of people from different continents. Also, certain spelling patterns need to be followed when translating foreign names into the Japanese language.

In this paper, we consider a method for using a neural network to solve the technical problems involved in using dictionaries.

6.4 Language Models with Neural Networks

6.4.1 Neural Probabilistic Language Model

The Neural Probabilistic Language Model was published by Bengio in 2003; this model makes predictions from the words that are already present [6-8]. This method maximizes the probability of the target word with the maximum likelihood principle in the score of the softmax function. When the word h that has already appeared is given, the probability that w_t appears is

$$\mathrm{P}(w_t|h) = \frac{\exp(\mathrm{score}(w_t, h))}{\sum_{all\ w'\ in\ dictionary} \exp(\mathrm{score}(w', h))}. \tag{6-6}$$

This equation is maximized by using the maximum likelihood method; it is the same as the following equation. The problem with this method is that the number of calculations increases with the increase in the size of the dictionary of member $\sum_{all\ w'\ in\ dictionary} \exp\left(\mathrm{score}(w', h)\right)$. In other words, J_{ML} can be written as follows:

$$\begin{aligned} J_{ML} =\ & \log \mathrm{P}(w_t|h) = \\ & = \mathrm{score}(w_t, h) - \log\left(\sum\nolimits_{all\ w'\ in\ dictionary} \exp\left(\mathrm{score}(w', h)\right)\right). \end{aligned} \tag{6-7}$$

6.4.2 Neural Probabilistic Language Model

Mikolov proposed the continuous bag-of-words (CBOW) method to speed up the train of the Neural Probabilistic Language Model and derive embedding vectors to improve the meaning better [6-8]. CBOW predicts w_t from the $2k$words $w_{t-k}, \dots, w_{t-2}, w_{t-1}, w_{t+1}, w_{t+2}, \dots, w_{t+k}$. We call this number $2k$ as the window size. However, the window size can be changed as a parameter. We will describe the experimental results by changing the window size in the later chapter. There are two aspects of the Neural Probabilistic Language Model. First, $\sum_{all\ w' \ in\ dictionary} \exp(\mathrm{score}(w', h))$ is calculated with not all the words in the dictionary but with randomly sampled words in the dictionary. This technique is called negative sampling. Second, each input word vector is compressed into the embedding vector, and all of these are added together. This method reduces the weight matrix to the output layer. An overview of CBOW is shown in Fig.6.7. It is known that the embedding vector derived by CBOW is a vector space based on the word meanings [6-6]. Even if the spelling of the word is different, if the surrounding words

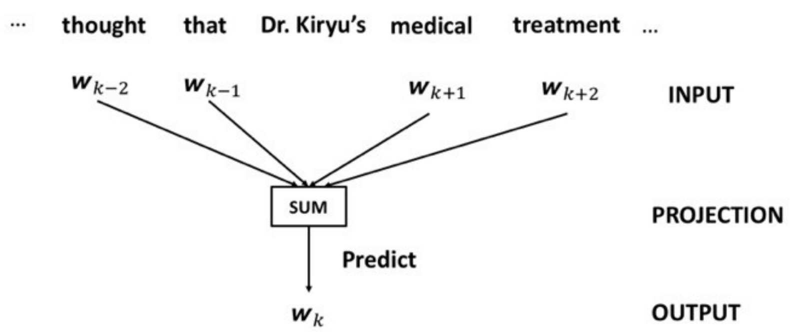


Fig. 6.7. Continuous Bag-of-Words

are similar, their embedding vectors will be similar (Fig.6.8)..

6.4.3 Long Short-Term Memory

Long Short-Term Memory (LSTM) is a kind of recurrent neural network (RNN). Adaptation to the language model was made by Mikolov et al. (2010) [6-9]. With RNN as the language model, continuous data ($\mathbf{w}_i$) is input and often handled in the task of predicting the next word. RNN has a feedback structure and calculates the output from the input (**w**i) and the feedback (**F**i-1). Various models have been proposed for this calculation method. However, we have used LSTM in this paper. The structure of the RNN (LSTM) model is shown in Fig.6.9.

Neural network A looks at the input *Wi* and outputs the value *Fi*. The loops allow information to

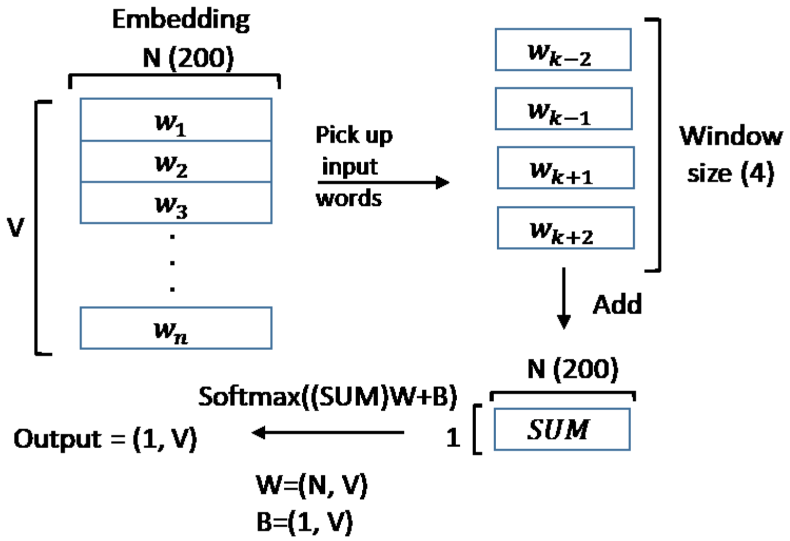


Fig. 6.8 Embedding vector derived by C-BOW

pass from one step of the network to the next. This is the mechanism of RNN (Recurrent Neural Network).

The network has an input layer x, hidden layer s (also called context layer or state), and the output layer y. Input to the network in time t is x(t), output is denoted as y(t), and s(t) is a state of the network (hidden layer). Input vector x(t) is formed by concatenating vector w representing a current word, and output from neurons in context layer s at time t − 1. Input, hidden and output layers are then computed as follows.

$$\mathrm{x(t)} = w(t) + s(t-1) \quad (6\text{-}8)$$

$$s_j(\mathrm{t}) = f\left(\sum_i x_i(t)u_{ij}\right) \quad (6\text{-}9)$$

$$y_k(t) = g\left(\sum_j s_j(t) v_{kj}\right) \tag{6-10}$$

where *f(z)* is sigmoid activation function:

$$f(z) = \frac{1}{1+e^{-z}} \tag{6-11}$$

g(z) is softmax function :

$$g(z_m) = \frac{e^{z_m}}{\sum_k e^{z_k}} \tag{6-12}$$

When the long time-series data coming into the RNN, the network becomes very deep in proportion to the time series length, and information is often not transmitted well. These are called vanishing gradients. LSTM has been used to solve this problem, which has three gates that update and control the cell states, these are the forget gate, input gate, and output gate.

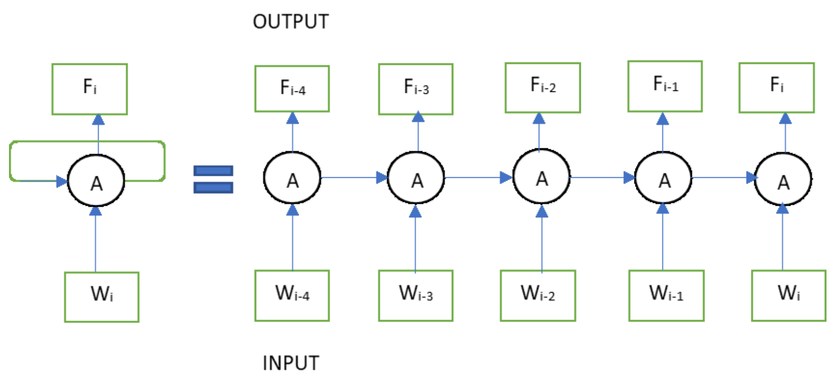


Fig. 6.9 Structure of RNN (LSTM)

The forget gate controls what information in the cell state to forget, given new information then entered the network. Fig.6.10

6.5 Prediction method using neural networks

6.5.1 Concepts used

As described in Section 6.4, we propose a method using a neural network without the use of proper noun dictionaries. A function is required to recognize the context and determine whether or not the target word needs to be converted to maintain confidentiality. LSTM is one of the neural network models and handles continuous data. LSTM is useful in the task of predicting the next word; therefore, we performed our experiment based on this model.

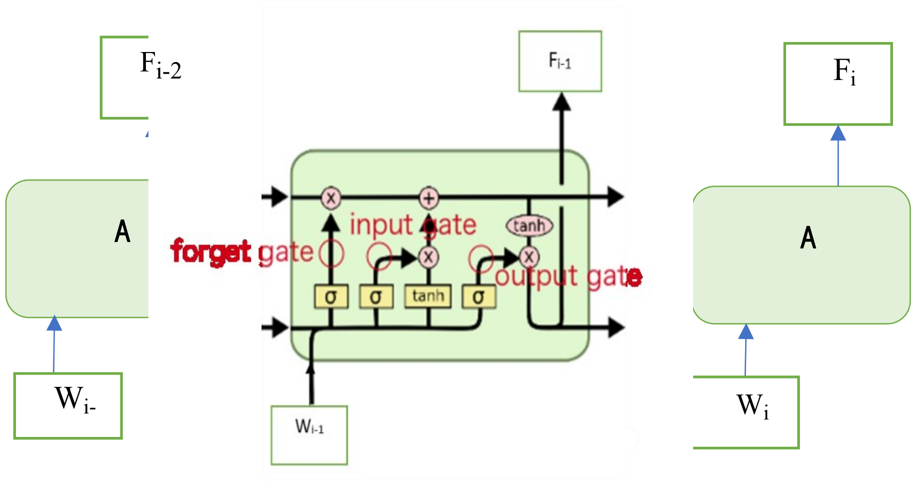


Fig. 6.10 Structure of LSTM

However, the goal of our research was a little different from the goal of the Neural Network Language Model (NNLM). In our study, we recognized a common concept that we should change words treated as having different meanings in the corpus (see Fig.6.7). In the previous tasks, the meanings of the words were used and recognized in the same context. For example, if the same predicted confidential words sometimes mean “names” and at other times means “place names”, they have completely different meanings and are returned as different letters of the alphabet. In other words, the concept of “confidential words” encompasses many words, and it will be difficult to derive this concept as an embedding vector. However, the CBOW model successfully expresses ambiguous meanings that were earlier difficult to express [6-9] [6-10]. Therefore, we decided to base this research on the CBOW model.

Our proposed approach is to predict confidential words from the words surrounding the target words. We assume that there are features in the distribution of words around confidential words. Therefore, the neural network model can capture the features of the distribution of words around confidential words. We experimented using CBOW to confirm our assumptions. The details of the experiment are described in Section 6.5.

The neural network predicts each word in the sentence from the words that have appeared so far.

In previous research, learning was performed using the word appearing in the sentence as the correct answer (For example, “dogs” and “cats” in Figure 6.11). However, in our research, a proper noun must be learned to predict it as a confidential word. In Fig.6.11, “Dr.Kiryu” is a proper noun; however, it must be predicted as A (confidential word).

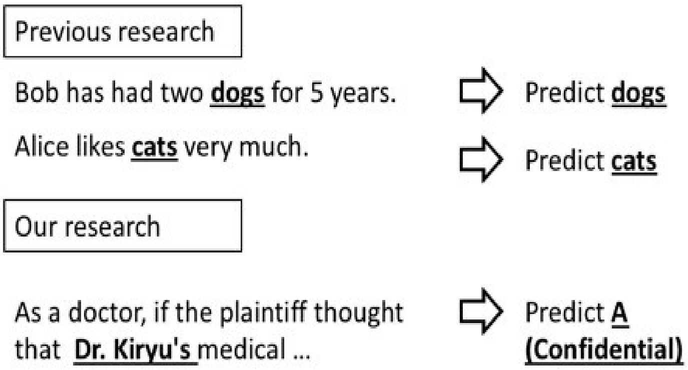


Fig.6.11 Difference between our research and previous research

6.5.2 Preliminary experiment

In this section, we explain the result of the experiment using the method given in Section 6.5.1. As mentioned in Section 6.5.1, a feature of the CBOW model is that the embedding vectors will be similar if the surrounding words are similar. Confidential words in the precedents published by the Japanese courts are usually converted to uppercase letters, such as A, B, and C.

The same letters cannot be used for different individuals in the same judicial precedent dataset.

$$\cos(\alpha, \beta) = \frac{\alpha \cdot \beta}{|\alpha||\beta|} \tag{6-13}$$

Therefore, if the CBOW model can capture the feature of "confidential words," the similarity of each of the converted confidential words (i.e., A, B, C, X, Y, Z) would also be high. In this paper, the similarity is defined as the cosine similarity, as shown in (6-13). Here α, and β are the embedding vectors of the words to be compared. The closer the cosine similarity is to 1, the higher is the similarity between the words. The judicial precedents dataset used in this experiment were 20,000 precedents available on the Japanese court website. The various parameters are shown in Table 6.1. Table 6.1 summarizes the results of calculating the cosine similarities of each confidential word by using the training result.

Table 6.1 CBOW parameters

Embeddings_Size	**200**
Batch_Size	**200**
Window_Size	**10**
Min_Learning_Rate	**0.001**
Min_count_vocab	**5**

The precedents are written in Japanese; therefore, very few are capitalized. It is more common for English words to have the first letter capitalized than Japanese words. In other words, the judicial precedents were Japanese sentences; therefore, it was extremely rare that an uppercase letter was used for English words. Table 6.2 shows the cosine similarity of the top 10 words to the confidential words (appearing as uppercase letters); these are the training results in precedents available for public viewing on the website of the Japanese court. As a result, the top 10 confidential words become uppercase letters.

From the above, we can see that the CBOW model can capture a part of the features of the distribution around the confidential words. Also, a previous study uses the CBOW model as a

Table 6.2 Top 10 Words Similar to Converted Words

Similarity to A		Similarity to B		Similarity to C		Similarity to D		Similarity to E		Similarity to F	
B	0.9352	A	0.9352	D	0.9251	E	0.9483	F	0.9487	G	0.9500
C	0.8833	C	0.9176	B	0.9176	F	0.9345	D	0.9483	E	0.9487
D	0.8071	D	0.8529	E	0.9165	C	0.9251	G	0.9345	H	0.9391
E	0.7612	E	0.8118	F	0.8893	G	0.8743	C	0.9165	D	0.9345
X	0.7363	F	0.7656	A	0.8833	B	0.8529	H	0.8806	I	0.9016
F	0.7094	G	0.7259	G	0.8463	H	0.8316	J	0.8328	C	0.8893
G	0.6772	H	0.7036	H	0.8186	A	0.8071	I	0.8326	J	0.8801
H	0.6517	X	0.6793	I	0.7550	I	0.7835	B	0.8118	K	0.8638
Y	0.6471	Y	0.6690	J	0.7521	J	0.7711	K	0.8003	M	0.8335
K	0.6032	K	0.6498	K	0.7358	K	0.7558	M	0.7743	L	0.8275

Similarity to G		Similarity to H		Similarity to I		Similarity to J		Similarity to K	
F	0.9500	I	0.9548	H	0.9548	I	0.9538	L	0.9589
H	0.9475	G	0.9475	J	0.9538	K	0.9527	M	0.9552
E	0.9345	J	0.9446	K	0.9435	H	0.9446	J	0.9527
I	0.9258	F	0.9391	L	0.9275	L	0.9429	I	0.9435
J	0.9241	K	0.9325	G	0.9258	M	0.9331	H	0.9325
K	0.8845	M	0.9054	M	0.9179	G	0.9241	N	0.9151
D	0.8743	L	0.9015	F	0.9016	N	0.8936	G	0.8845
M	0.8653	E	0.8806	N	0.8751	F	0.8801	O	0.8787
L	0.8569	N	0.8632	O	0.8346	O	0.8717	F	0.8638
C	0.8463	O	0.8356	E	0.8326	R	0.8367	Q	0.8614

Similarity to L		Similarity to M		Similarity to N		Similarity to O		Similarity to P	
K	0.9589	L	0.9583	M	0.9508	P	0.9381	O	0.9381
M	0.9583	K	0.9552	L	0.9358	N	0.9225	Q	0.9192
J	0.9429	N	0.9508	O	0.9225	M	0.9221	R	0.9026
N	0.9358	J	0.9331	K	0.9151	Q	0.9128	N	0.8864
I	0.9275	O	0.9221	Q	0.9119	R	0.9096	S	0.8789
H	0.9015	I	0.9179	R	0.9054	L	0.8995	M	0.8683
O	0.8995	H	0.9054	J	0.8936	K	0.8787	L	0.8632
Q	0.8862	Q	0.8973	P	0.8864	J	0.8717	T	0.8449
R	0.8754	R	0.8915	I	0.8751	S	0.8604	K	0.8397
P	0.8632	P	0.8683	S	0.8730	W	0.8419	J	0.8265

Similarity to Q		Similarity to R		Similarity to S		Similarity to T		Similarity to U	
R	0.9212	Q	0.9212	T	0.9296	S	0.9296	W	0.9251
P	0.9192	S	0.9136	R	0.9136	U	0.8977	S	0.9044
O	0.9128	O	0.9096	U	0.9044	R	0.8805	T	0.8977
N	0.9119	N	0.9054	Q	0.9020	W	0.8609	R	0.8947
S	0.9020	P	0.9026	P	0.8789	Q	0.8592	Q	0.8928
M	0.8973	U	0.8947	N	0.8730	N	0.8460	N	0.8624
U	0.8928	M	0.8915	O	0.8604	P	0.8449	M	0.8534
W	0.8883	T	0.8805	M	0.8541	O	0.8418	O	0.8397
L	0.8862	W	0.8774	W	0.8524	M	0.8313	L	0.8390
K	0.8614	L	0.8754	L	0.8424	L	0.8255	V	0.8234

Similarity to V		Similarity to W		Similarity to X		Similarity to Y		Similarity to Z	
W	0.8423	U	0.9251	A	0.7363	Z	0.8119	Y	0.8119
M	0.8271	Q	0.8883	Y	0.7019	X	0.7019	V	0.7378
U	0.8234	R	0.8774	Z	0.6883	P	0.6780	W	0.7361
N	0.8173	N	0.8613	P	0.6807	B	0.6690	Q	0.7321
R	0.8046	T	0.8609	B	0.6793	A	0.6471	O	0.7205
S	0.7919	S	0.8524	C	0.6753	O	0.6453	U	0.7104
Q	0.7891	M	0.8479	V	0.6510	Q	0.6320	P	0.7101
O	0.7887	V	0.8423	O	0.6264	C	0.6315	N	0.7067
T	0.7713	O	0.8419	K	0.6191	W	0.6034	M	0.7065
K	0.7702	P	0.8226	D	0.6180	M	0.6029	K	0.6921

predictor based on the meanings of words [6-11]. Therefore, in this paper, we use several neural networks based on the CBOW model to predict confidential words and consider a network model effective for predicting them.

6.6 Predicting confidential words

In this section, we describe an experiment to predict confidential words by using neural networks.

I will explain the prediction mechanism of the confidential words (see Fig.6.12)

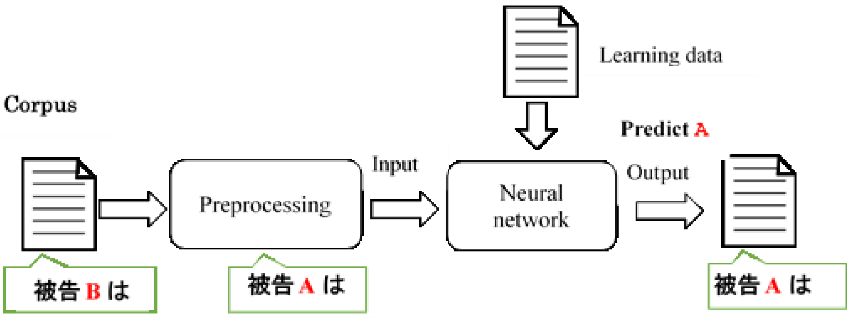


Fig.6.12. Prediction mechanism of the confidential words

At first, we converted the confidential words contained in the datasets to the uppercase letter "A" and separated the Japanese words with spaces by using MeCab, a Japanese morphological analyzer in the preprocessing. When the Japanese precedents (corpus) contained the confidential words replaced by "A" are entered into the neural network, they are learned by the neural network, predict the confidential words.

We propose two models: one model imitates a human being, (6.6.1 (1)) and the other model is based on the concept of the CBOW model (6.6.1(2)).

6.6.1 Proposed Model

(1) Bi-directional LSTM LR

Bidirectional recurrent neural networks (RNN) are just putting two independent RNNs together.

The input sequence is fed in normal time order for one network, and in reverse time order for another putting two independent RNNs together.

The outputs of the two networks are usually concatenated at each time step. This structure allows the networks to have both backward and forward information about the sequence at every time step. Fig.6.13 shows a structure of Bi-directional LSTM.

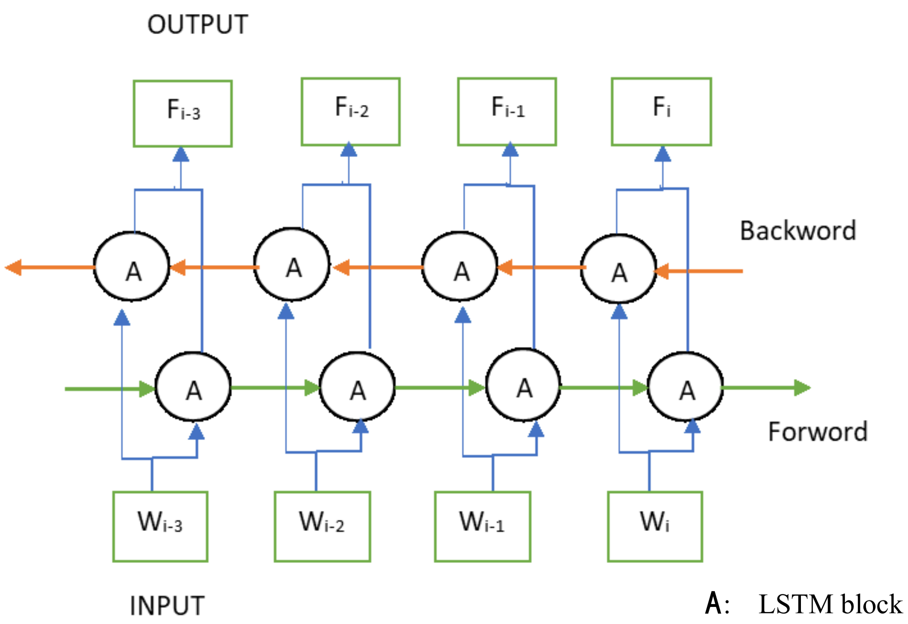


Fig.6.13 Structure of Bi-directional LSTM

Bi-directional LSTM LR is a model that imitates the anonymization done by humans. When humans perform anonymization, they make a judgment after reading to the left and the right of the target word. Therefore, it becomes a shape as shown in Fig.6.14 (c). The input order on the back (right side) of the target word is the reverse of the sentence order because we assume that the words closer to the target word have higher importance.

(2) Sum-LSTM

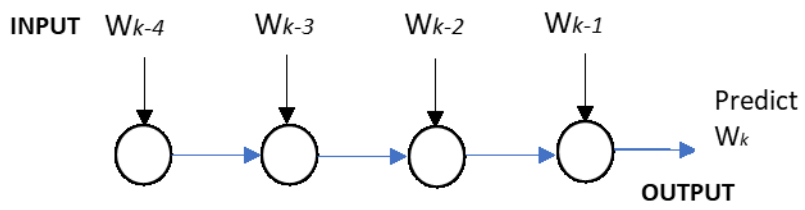


Fig.6.14(a)　Simple

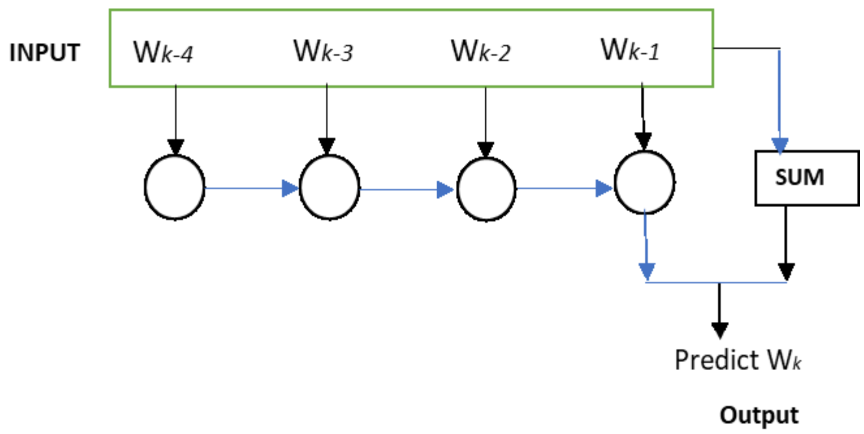


Fig.6.14(ｂ)　Sum LSTM

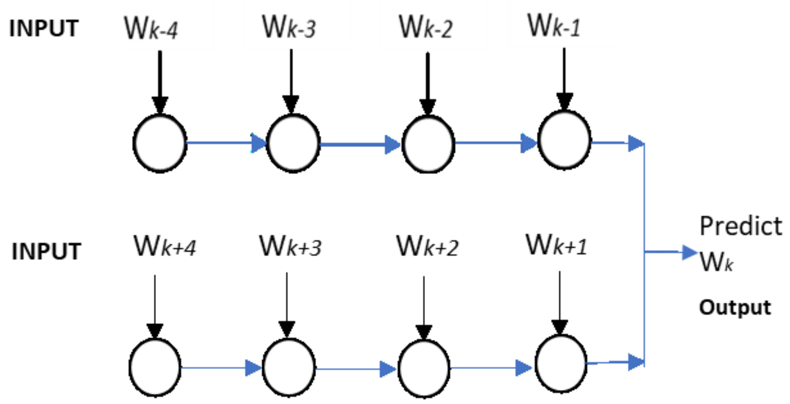


Fig.6.14(c)　Bi-directional　LSTM LR

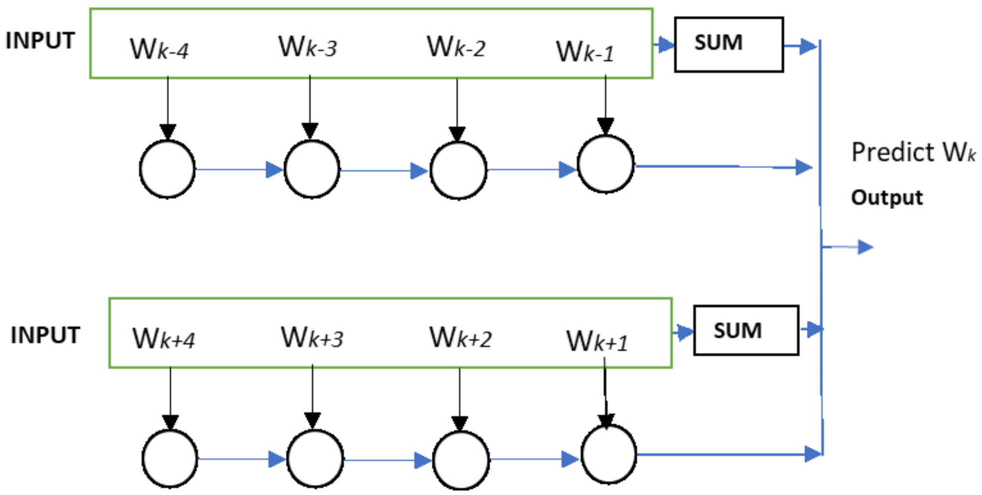


Fig.6.14(d) Sum Bi-directional LSTM LR

The Sum-LSTM, which is based on the CBOW model, is a model that validates the effectiveness of the model (see Fig. 6.14 (b)). In addition to the normal LSTM calculation, the total of all the input vectors is calculated and activated by the softmax function in the output layer. The model combined with the Bi-directional LSTM LR model is shown in Fig. 6.14 (d).

And we proposed the simple LSTM model as shown in Fig.6.14 (a).

6.6.2 Corpus and Evaluation method

We used 50,000 judicial precedents for the training data and 10,000 judicial precedents for the test data. These data included the records of trials from 1993 to 2017. We used the precedent database provided by TKC, a Japanese corporation [6-12].

Various parameters are shown in Table 6.3.

Window size means chunk size, describe the input words size before or after the target word. Fig. 6.14 shows the window size is 4 to explain the model, but in this experiment, it is 10. We converted all the confidential words into the uppercase letter “A” and separated the Japanese words with spaces by using MeCab, a Japanese morphological analyzer.

. MeCab was required because we were using the Japanese judicial precedents dataset [6-13].

Table 6.3. Parameters of the model

Embedding size	200
Window size	10
Batch size	200
Learning rate	0.001
Forget bias	1.0
Loss function	Softmax entropy
Optimizer	Adam

Also, word prediction required stop words; therefore, word prediction was not excluded in this experiment.

6.7 Results of the Experiment

In this experiment, we also prepared a simple LSTM model to compare the two models proposed in Section 6.6.1. This model had a three-layered structure: an input layer, a hidden (LSTM) layer, and an output layer. The size of the hidden units was 200. The input/output layer size was the same as the vocabulary size (approximately 200,000 in our corpus).

The Bi-directional LSTM LR model had two simple LSTM model structures, and Sum-LSTM also inherited the simple LSTM structure. Besides, we combined the Sum-LSTM and LSTM LR models and named it Sum-Bi-directional LSTM LR (see Fig. 6.14(d)).

This experiment was conducted using the four models shown in Fig.6.14. Also, the embedding vector was 200 for all models [6-14].

For accuracy, we used perplexity (PPL) that was used in previous research for predicting the next word.

In general, perplexity is a measurement of how well a probability model predicts a sample. In the context of Natural Language Processing, perplexity is one way to evaluate language models. In machine learning, the term perplexity has three closely related meanings. Perplexity is a measure of how easy a probability distribution is to predict. Perplexity is a measure of how variable a prediction model is. And perplexity is a measure of prediction error. The third meaning of perplexity is calculated slightly differently but all three have the same fundamental idea

PPL was given by the following equation:

$$PPL = 2^{-\frac{1}{N}\sum_{i=1}^{N} log_2\, p(w_i)} \qquad (6-14)$$

In (6-14), $P(w_i)$ is the probability, and N indicates the total number of the words. PPL represents the number of prediction choices that are narrowed down to neural networks. The smaller the value, the better the prediction results.

Table 6.3 shows the results of the experiment using the proposed neural network. CW_PPL is

Table 6.3 Experimental results of proposed neural networks

	Simple-LSTM	Bi-directional LSTM LR	Sum-LSTM	Sum-Bi-directional LSTM LR
PPL	4.851	4.656	4.603	**4.462**
CW_PPL	56.277	**37.343**	72.463	77.031

the average PPL of the test data whose answer reflects the confidential words. However, PPL is the average of all the test data. The PPL scores in Table 6.3 show that the Bi-directional LSTM LR model decreased by 0.195 as compared with the Simple-LSTM model. Also, the Sum-LSTM model decreased by 0.247 as compared with the Simple-LSTM model. The combination of the two methods scored the best results, which was 4.462. Therefore, the proposed models are effective for PPL in our corpus.

However, if we look at the results of CW_PPL directly related to the task of predicting the confidential words, we will find that the difference of scores is at least 32.492 between PPL and CW_PPL. This result suggests that the task of predicting confidential words is more difficult than the task of predicting other words. Also, each CW_PPL score shows that the Bi-directional LSTM LR model decreased by 18.934 as compared with the Simple-LSTM model. (The score for the Bi-directional LSTM LR model was 37.343, which was the best score). However, the Sum-LSTM model increased by 16.186 as compared with it. Furthermore, the combination of the two methods recorded the worst score.

In PPL, we found that all the proposed methods were more effective than the simple model. However, for the prediction of confidential words, only the Bi-directional LSTM LR model showed good results. Sum-LSTM based on CBOW might have produced these results. CBOW is an effective model for paraphrasing words, and Sum-LSTM also uses this mechanism. Therefore, when Sum-LSTM predicted a word whose answer is "confidential," the CW_PPL became worse because there was a possibility of paraphrasing words such as "plaintiff," "defendant," "doctor," and "teacher." Knowing the paraphrased words of the confidential words meant that the embedding vectors of the confidential words could be successfully generated. This meant that the model could recognize the meaning of "confidential." However, the prediction accuracy did not improve; therefore, there was a problem in calculating the probability of the prediction task. To solve this problem, we could exclude these paraphrasable words from the choices when calculating the probability. It is also important to examine scores other than PPL.

6.8 Improving accuracy to predict confidential words in judicial precedents using the neural network

In the conclusion of the first experiment, we got the achievement that our proposed model using a neural network was effective for predicting confidential words. Nevertheless, these models were necessary to improve accuracy. [6-15]

In this section, we reviewed the results of the previous experiment, then we experimented with the proposed neural network by changing the window size and chose the proper value. After that, we would enhance the pre-process of input datasets. This operation is important for learning by the neural network.

6.8.1 The experiment with window size changed

(1) Problem of the previous experiment

We had the following problems through the first experiment in the case of the Bi-directional LSTM model that was the best score compared with other models.

・The score was bad with the continuous numeric words. It was about ten times worse than the target words. We suppose that our neural network model had a weakness for the continuous numeric words to train the many test dataset involved numeric words.

・CW_PPL became worse because there was a possibility of paraphrasing words such as "plaintiff," "defendant," "doctor," and "teacher".

To solve this problem, we could exclude these paraphrasing words from the choices when calculating the probability. It is also important to examine scores other than PPL.

These models were based on Mikolov's paper [6-9]. However, there were other important parameters as well.

(2) The aim of the experiment and plan to improve accuracy

Our ultimate goal is to obtain high accuracy when automatically predicting confidential words. Therefore, we solved the problem above described and upgraded the detection accuracy for confidential words using a neural network. We designed an experiment based on the first experiment, that is, the Bi-directional LSTM-LR model (Fig.6.15).

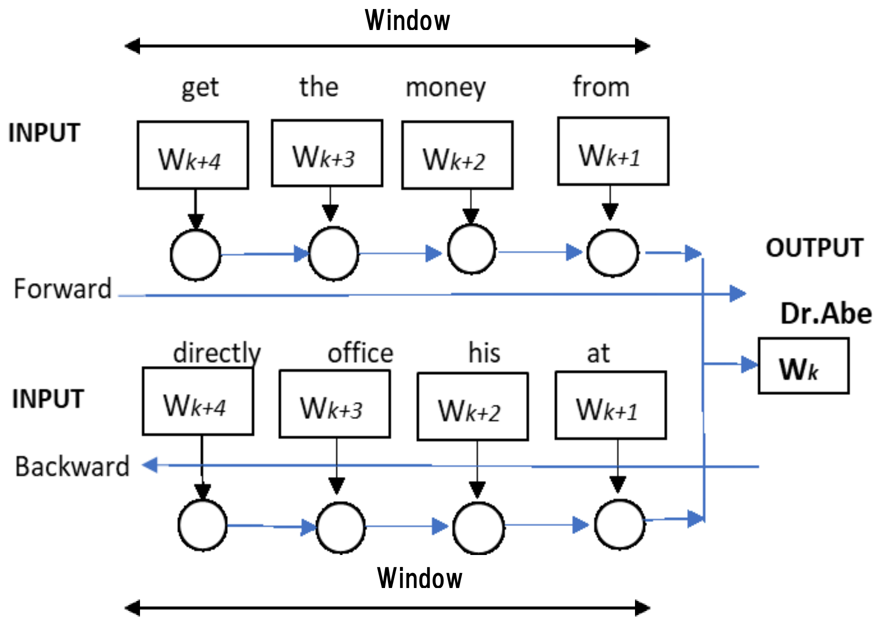


Fig.6.15. Experiment using the model (Bi-directional LSTM LR)

First, we review the parameters, such as the window size, of our neural network model to obtain the proper value for the window size. Besides, before learning step in our proposed neural network, to predict the target words more easily, we enhance the input dataset via preprocessing by changing to a powerful morphological analyzer. Simultaneously, we avoid unnecessary words, i.e., “stop words”, in the input dataset.

(3) Changing the size of window

First, we focused on the window size. In general, a neural network such as LSTM has a chunk size (which we call the window size in this paper). The chunk size is the number of (output) frames for each chunk of data evaluated in the training or decoding.

RNN or LSTM models, or "chain" models, are always trained on fairly large chunks (generally in the range of 40–150 frames). During decoding, we also generally evaluate the neural network using fairly large chunks of data (e.g. 30, 50, or 100 frames);

This is usually referred to as the frames-per-chunk. For recurrent networks, the chunk-size/frames-per-chunk and the extra-left-context and extra-right-context are approximately the same during training and decoding because this generally gives the best results (even though sometimes it is the best to make the extra-context values slightly larger in decoding).

One might expect that in decoding time longer context would always be better; however, this does not always seem to be the case [6-16]. If the length of a chunk is too short, the association between chunks cannot be learned, which will affect the accuracy of the LSTM model [6-17].

In our first experiment, we used a window size at 10 frames (words), as shown in Table 6-1. At that time, we did know if this number was appropriate.

Accordingly, we changed the "window size" parameter from 10 to 20 because we expected the PPL score to be improved by larger training samples. We expected to obtain the correct results [6-18]. Besides, we changed the window size from 10 to 5 to investigate how the training ability is affected by narrowing the window size. We see that the PPL values significantly worsen in this case. The PPL scores are seven times larger than those in the original experiment, and the CW_PPL values further worsened. Therefore, the window size greatly influences the accuracy when detecting confidential words. Next, we enlarged the window size from 10 to 20 or 30 and trained the neural network. The resulting PPL and CW_PPL scores are shown in Fig. 6.16. The PPL and CW_PPL scores are best when the window size is 20. When we enlarge the window size to 30, the PPL score is good but the CW_PPL score is worse. The CW_PPL score is likely poor when the window size is too large because, in this case, a large number of "confidential words" are contained in the window and the neural network has many choices.

We also changed the value of "epoch". When training a neural network, one epoch indicates one pass through the full training set. Usually, a few iterations are used. We experimented with epoch values of 1, 5, and 100. Normally, when the epoch value is larger, the predicting accuracy is better. However, if the epoch value is too large, overlearning will occur and the accuracy will decrease. In our experiment, even when the epoch value was 1, PPL changed little while CW_PPL worsened. We assume that an epoch value of 100 is better for predicting confidential

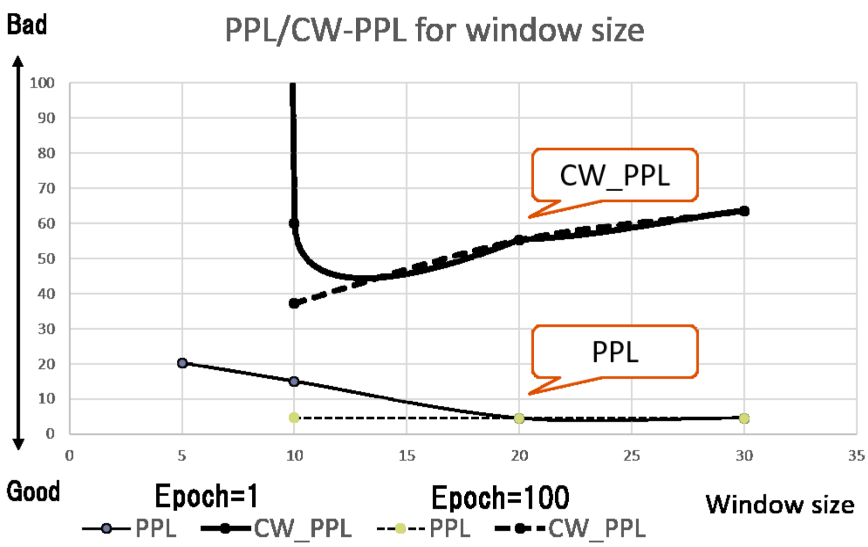


Fig.6.13. PPL/CW_PPL for window

words. The result is shown in Fig.6.13.

(4) Enhancement of the input dataset processing before learning

① Update of the morphological analyzer "MeCab" to "MeCab-ipadic- NEologd"

all the confidential words contained in the datasets are converted to the uppercase letter "A" and separated the Japanese words with spaces using "MeCab", a Japanese morphological analyzer [6-19] because we were using the Japanese judicial precedent dataset.

These processes were performed in the data-preprocessing step (Fig. 6.17).

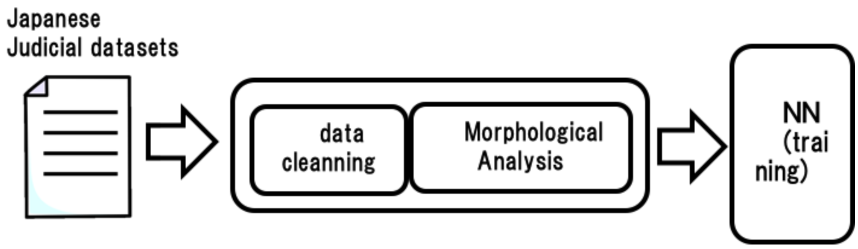


Fig.6.17 Data preprocessing step.

In the second experiment, we decided to use "MeCab-ipadic-NEologd", which is a more powerful morphological analyzer, has better performance than "MeCab", and is open software. We expected the prediction accuracy would be improved for the confidential words (Fig. 6.18).

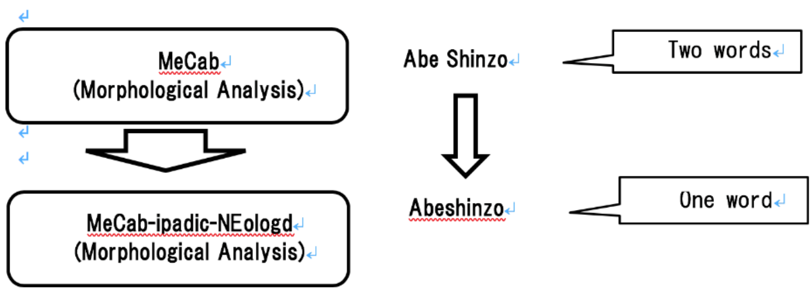


Fig.6.18. MeCab upgrade

However, the result of the experiment shows that the accuracy decreased compared to the first experiment (Table 6.4). That is, the PPL value didn't change but the CW-PPL value became two times worse. It is possible that, because the ability of the morphological analysis increased, overlearning occurred. Also, the neural network parameters changed. Therefore, we need to continue the experiment to obtain the proper parameters, e.g., the epoch size and the window size.

② Exclude Legal terminology as stop words

Because we are dealing with judicial words in the corpus, legal terminologies frequently appear in the corpus. These words do not necessarily predict confidential words. Therefore, it is important to exclude these legal terminologies from the corpus. Accordingly, we registered these judicial words in a stop words list. Then, we removed these words from the corpus. The evaluation of the results will require further study.

Table 6.4. Results in the case of using MeCab-ipadic-Neologd

MeCab	MeCab-ipadic-NEologd		MeCab	
Window size	**10**	**20**	10	**20**
PPL	**5.81**	**4.78**	4.656	**4.53**
CW_ PPL	**83.8**	**85.56**	**37.343**	55.34

Chapter 7

Predict the confidential words by POS tag

7.1 Introduction

As we couldn't the good accuracy for predicting the confidential word, we investigated the algorithm of our model and reviewed them. Generally, words have meanings and parts of speech (POS) in the dictionary. We found almost all of the confidential words have a proper noun as POS. So, we considered if a POS tag is added to the neural network with words, it may be possible to learn better and improve accuracy. We proposed a new prediction mode using the neural network

Table.7.1. Comparison of Japanese morphological analyzer

Morphological analyzer	Application example
MeCab	Fastest speed and most commonly used
JUMAN	KNP available
JUMAN++	High accuracy but slow speed
Sudachi	Used for retrieve system

combined with a POS tag that is extracted by Japanese morphological analyzer such as “MeCab”, “JUMAN”, “JUMAN++”, and “Sudachi”. This morphological analyzer is compared in Table 7.1.

We adopted “MeCab” as a morphological analyzer of our new model because it is the most commonly used and fastest speed.

7.2 MeCab as CRF

“Mecab” is the most powerful tool to extract the POS tag from words in Japanese precedents. It is a well-known Japanese morphological analyzer [7-1]. CRF (Conditional Random Field) is a successful named entity extraction output technique to label information such as POS tagging. Japanese sentences have no spaces between words, so “Mecab” insert a space between each word and tag it (POS). If the word is ‘A', it is a confidential word. The POS corresponding to ‘A’ is replaced by the proper noun as described in Fig.7.1 and section 7.3.

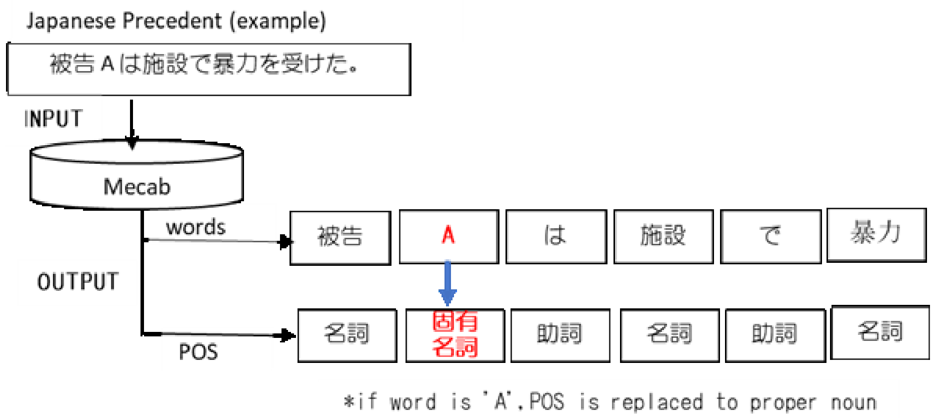


Fig.7.1 POS tagging using Mecab

7.3 The Bi-directional LSTM-LR combined the POS tag (new proposed model)

In the previous model, the corpus (words) were input to the neural network like natural language processing. To improve the CW_PPL score, we attempted to input the POS tag corresponding to the word extracted by "Mecab" (CRF) to the previous model (Bi-directional LSTM-LR). The outline of this proposed model is shown in Fig.7.2. It is different from conventional natural language processing technology. Each word (W_ki) describes a word, and each Other or Noun is the

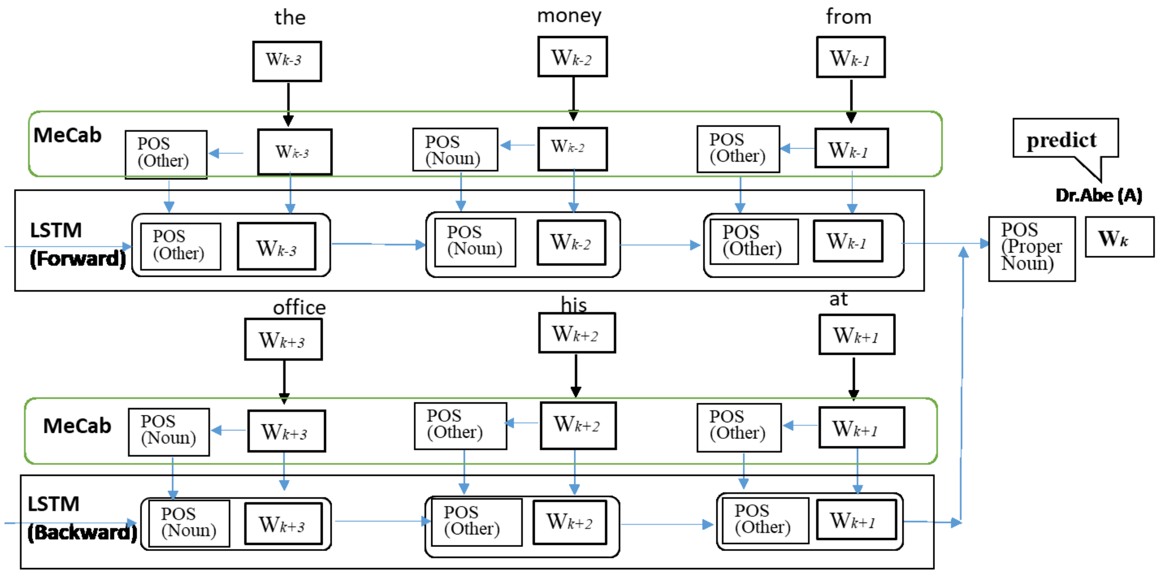


Fig.7.2 A proposed model combined with the POS tag

associated POS of the word (W_{ki}).

When the Japanese precedents (corpus) is input into "MeCab," "MeCab" separates the words with space and tag them (POS). If the confidential word (A) appears in the precedents, proper noun, e.g., the place name or the personal name is tagging to the confidential word. Next, we assign a unique index to the POS information and make a part dictionary, merging them into the input data (word) for the new neural network via an embedding vector.

Then preprocessed words, which the space inserted among each word, and POS tags by MeCab are input to the neural network that is Bi-directional LSTM.

7.3.1 Prediction method of the new proposed model combined the POS tag

(1) Preprocessing

Before the original corpus (Japanese precedent) is input to the neural network, they are preprocessed as described in section 6.6. Characters and words that are not relevant to learning by the neural network are replaced with blanks. This is a cleaning work described in Fig.7.3. Then, the preprocessed corpus is separated into the words with space and tag them (POS) by "MeCab". Further, it is normalized the character code to prevent garbled characters (word normalization). Next, the words which are not relevant to learning are omitted or replaced with a blank.

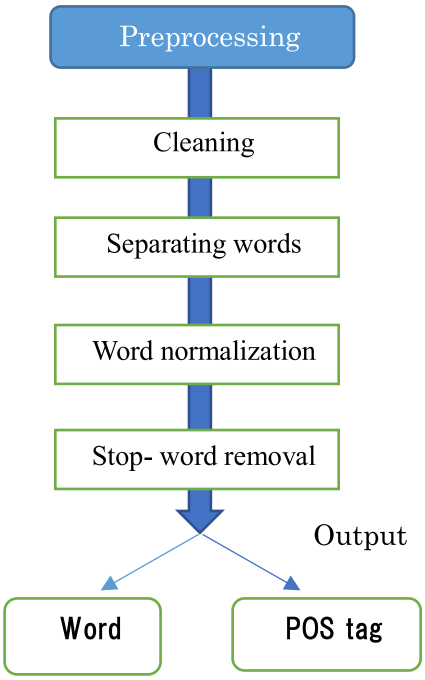


Fig.7.3 Preprocessing of the corpus

Our study doesn't yet have this feature. At last, these preprocessed words and POS tags are output.

If the confidential word (A) appears in the precedents, proper noun, e.g., the place name or the personal name is tagging to the confidential word (see. Fig.7.1).

We describe the preprocessing algorithm by MeCab as bellows.

If x is sequence and its y is label sequence, the probability ($p(y \mid x)$) assigned to a label sequence for a particular sequence of characters by a MeCab (CRF) is given by the - equation below [7-2] :

Input corpus $\boldsymbol{D} = \{(x_1, y_1), (x_2, y_2), \cdots\cdots, (x_n, y_n)\}$, (7-1)

$$p(y|x) = \frac{1}{Z(x,w)} \exp(w\emptyset(x,y)) \qquad (7\text{-}2)$$

$$Z(x,w) = \sum_y \exp(w\emptyset(x,y)) \qquad (7\text{-}3)$$

$Z(x,w)$ is a normalization term, w is weight parameter , Φ is feature function.

Output $\boldsymbol{y}$

$$\boldsymbol{y} = \arg\max_{y} \frac{1}{Z(w\emptyset)} \exp(w\emptyset(x,y) \qquad (7\text{-}4)$$

The output $\boldsymbol{y}$ is a POS tag.

Then preprocessed words, which the space inserted among each word, and POS tags by MeCab are input to the neural network that is Bi-directional LSTM.

Next, we assign a unique index to the words and POS tag. Then, we create a word dictionary and a POS dictionary respectively, concatenating them into the new neural network via an embedding vector. They are added both forward LSTM and backward LSTM of the target word. The input order on the back (right side) of the target word is the reverse of the sentence order because we assume that the words closer to the target word have higher importance.

The identification mechanism is shown in Fig. 7.4. The summary of the algorithm is described as bellows and detail is given in the appendix [7-3].

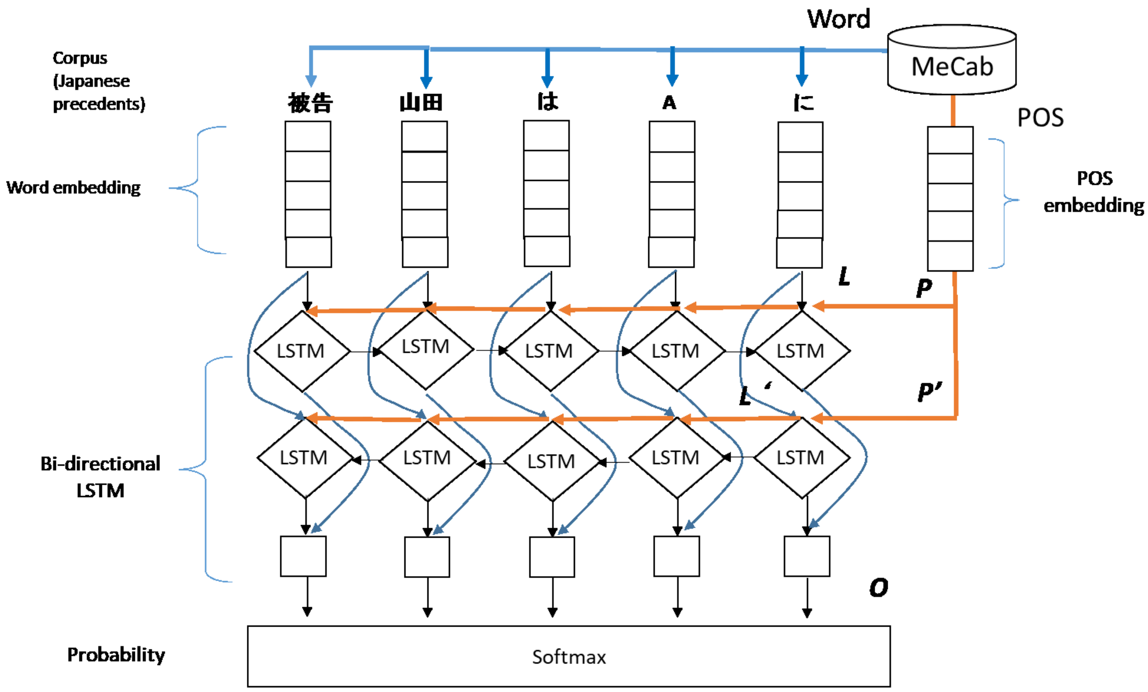


Fig.7.4 Identification mechanism of the proposed model.

Input data

$\boldsymbol{L} = (\mathrm{w_i})_{\mathrm{i}}^{\mathrm{bj}}$: $w_1, w_2, \cdots w_{10}$: word(backward) (7-5)

$\boldsymbol{L'} = (\mathrm{w_i})_i^{fj}$: $w_{-1}, w_{-2}, \cdots w_{-10}$: word(forward) (7-6)

$\boldsymbol{P} = (p)_i^{bj}$: $p_1, p_2, \cdots p_{10}$: POS(backward) (7-7)

$\boldsymbol{P'} = (p_i)_i^{fj}$: $p_{-1}, p_{-2}, \cdots p_{-10}$: POS(forward) (7-8)

Output

$$O_i = \text{LSTM}\,(\boldsymbol{L} + \boldsymbol{L'} + \boldsymbol{P} + \boldsymbol{P'}) \qquad \textbf{: \textit{output}} \qquad (7\text{-}9)$$

7.4 The experiment of the new proposed model combined the POS tag

We experimented using the proposed model combined with the CRF. We used 10,000 judicial precedents for training data and 5,000 judicial precedents for test data from 2013 to 2017. The data is as same as the previous one. However, the number of training data was smaller because the data is two times larger than the previous one by adding POS information. Various parameters are shown in Table 7.1. The evaluation method is also the same as the previous one.

Table 7.1 Parameters of the model

Hidden layer	100
Embedding size	200
Window size	10
Batch size	200
Learning rate	0.001
Loss function	Softmax function
Optimizer	Adam

Table 7.2. The result of the second experiment

	proposed model	Bi-directional LSTM-LR
PPL	4.1	5.2
CW_PPL	**28.6**	**40.7**

Results of the experiment compared with the Bi-directional LSTM-LR mode using the same input corpus are shown in Table 7.2.

We got the PPL score was 23% improvement in accuracy over the previous model (Bi-directional LSTM-LR), and the CW_PPL score also was a 30% improvement inaccuracy. Therefore, we found the Bi-directional LSTM-LR model combining words and POS was very effective in predicting the confidential words. However, the CW_PPL score needs further improvement.

7.5 Improve the preprocessing algorithm

Before learning the input corpus by the neural network, the preprocessing of the corpus is necessary. For example, many punctuation marks, e.g., "「」" and "（）", those mean separators often appear in Japanese precedents. As they can only be noise for learning, we replaced them with brank (Fig.7.5). However, the punctuation mark "。" is not omitted because to prevent the flow of the sentence. This preprocessing is also done in the first experiment. In Japanese precedents, confidential words are replaced by not only half-width uppercase alphabets but also by full-width alphabets in uppercase and lowercase. In the previous algorithm, when only a single half-width capital letter of the alphabet appeared in the precedent, we replaced it with a half-width uppercase letter "A" so far as shown in Fig.7.6 (above). Therefore, we improve this algorithm that if both single half-width and full-width alphabet appear in the precedent, we replaced it with a half-width uppercase capital letter "A" as the confidential word as shown in Fig.7.6(below).

7.6 Results of the experiment after the improved algorithm

Table 7.3 shows the results of the experiment after the improved algorithm. CW_PPL shows that the proposed model after the improved algorithm decreased by 35.6 compared with the previous model (Bi-directional model). It was significantly better than the previous model. Finally, we got the CW_PPL score was 88% improved in accuracy and 20 % improvement for detecting the target word (PPL) compared with the previous model.

As a result, we confirmed our proposed model (Bi-directional LSTM-LR combined with POS tag) had high accuracy for predicting the confidential words.

As we got an excellent predicting ability with our proposed model, we need to confirm if it is

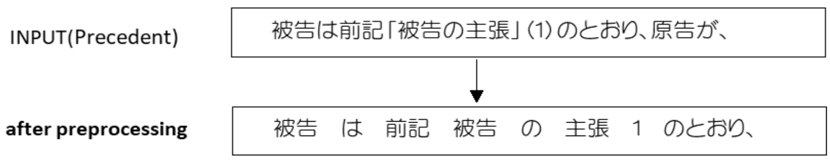


Fig.7.5. Example of the Preprocessing of Japanese

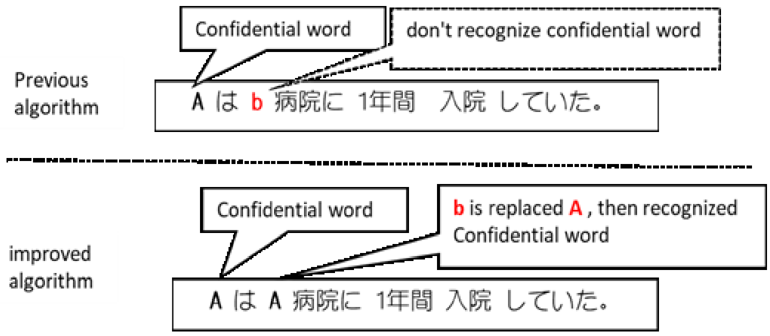


Fig.7.6 Example of the improved algorithm

Table 7.3 **R**esults of the experiment after the improved algorithm

	New proposed model after the improved algorithm	Previous model (Bi-directional LSTM-LR)
PPL	**4.1**	5.2
CW_PPL	**5.1**	40.7

practical or not in the next step.

7.7 Extension of the proposed model to increase confidential word detection accuracy

In an actual legal record, evaluating whether a confidential word is correctly recognized or falsely identified is essential. Therefore, we used the evaluation parameters of "recall," "precision," and "F1" to confirm the possibility of practical use. Defining the threshold value of CW_PPL, which determines whether the confidential word is truly recognized, is necessary [7-4].

7.7.1 Evaluation parameters

We then evaluated how our proposed model affected some types of anonymized precedents. We used the following parameters to examine accuracy.

$$\text{Recall} = \frac{TP}{TP + FN} \tag{7-10}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{7-11}$$

$$\text{F1} = \frac{2Recall \times Precision}{Recall + Precision} \tag{7-12}$$

TP: a true positive (i.e., when the actual word was confidential word, predicted it correctly)

TN: a true negative (i.e., when the actual word was not confidential word, but predicted it incorrectly)

FN: a false negative (i.e., when the actual word was confidential word, but predicted it incorrectly)

FP: a false positive (i.e., when the actual word was not confidential word, predicted it correctly)

The actual positive is "TP + FN," and the total predicted positive is "TP + FP."

"Recall" is an index indicating the fraction that was correctly predicted among all the positive words. "Precision" is an index that shows the fraction of positive words among all the words that were predicted to be positive. F1 is an index that balances recall and precision.

7.7.2 Determining the threshold value of CW_PPL

If the CW_PPL of a confidential word ("A") was lower than 60 (threshold), we defined it as a correct word (TP). Otherwise, it was ignored (FN). A threshold of 60 was selected as it is the optimal value judged from the area under the curve (AUC)-receiver operating characteristic (ROC) curves (Fig. 7.5). In machine learning, performance measurement is an essential task. In classification problems, we adopt the AUC-ROC curve [7-5], a key evaluation metric for checking any classification model's performance. The ROC curve is a performance measurement for classification problems at various threshold settings. While ROC is a probability curve, and AUC represents the degree or measure of separability that indicates how much the model is capable of distinguishing classes. The higher the AUC, the better the model is at predicting 0s as 0s and 1s as 1s. The ROC curve is plotted with a true positive rate (TPR) against the false positive rate (FPR), where TPR is on the y-axis, and FPR is on the x-axis, using the equations below.

$$\text{TPR (True Positive Rate)} = \frac{TP}{TP+FN} \qquad (7\text{-}13)$$

$$\text{FPR (False Positive rate)} = \frac{FP}{FP+TN} \qquad (7\text{-}14)$$

The curve is deemed better when it is located toward the upper left.

It is critical that TPR (=Recall) is large, and FPR is small. Therefore, we choose the threshold to be 60 that was the closest point of the ideal curve.

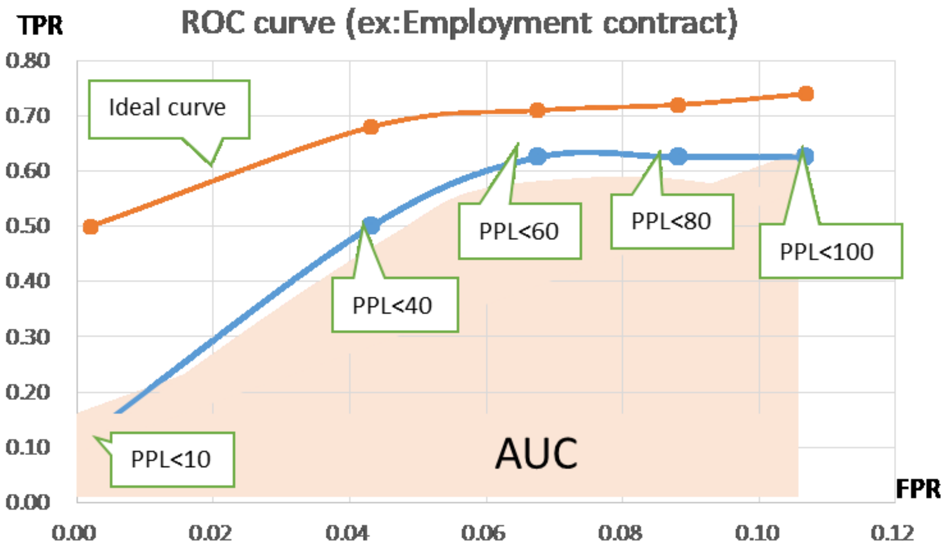


Fig. 7.5. ROC curve

7.8 Applying the proposed model to the plain precedents

7.8.1 Experimental results for anonymized precedents

We decided if the CW_PPL of the confidential word (A) was lower than 60 (it was threshold), we recognized it as the correct word (TP). If not, we couldn't recognize it (FN). The actual sample precedent of predicting the confidential word (A) was described in Fig.7.6.

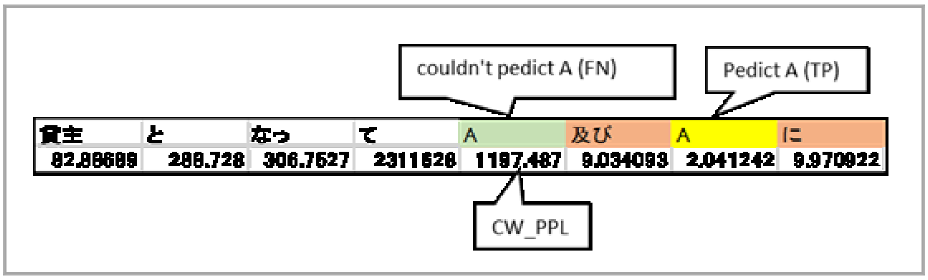


Fig.7.6. Predict A with actual sample precedent

The results of the experiment in various types of anonymized precedents are shown in Table

7.4. Besides, we describe an example of predicting the result in current precedents after anonymization in Fig.7.7.

As a result, the recall score showed good values for some types of precedents. In particular, when similar phrases appeared multiple times in some precedents, our proposed model could recognize the confidential word well. Also, if we had more total words in one precedent, the recall score was good ranging from 69% to 84%. The confidential words were sometimes replaced with non-English letters, such as the Greek letters γ or β; then, our proposed model could not predict them as confidential words. However, if they are replaced as “A” during preprocessing, the recall score would be good.

In the results, the recall rate was ≥80% in about 60% of the cases. However, the precision rate was low at approximately ≤30% and the F1 score was at ≤50%. In particular, when the number of appearances of confidential words was small, the matching rate became low. First, however, confidential words need to be properly recognized as true confidential words.

4	日	で	あり	、	被告	A	又は	被告	A	が	最後	に	本件	事故	による	損害	の	填補	として
458757.6	114517.1	11704.89	269.8959	5734.829	136.2987	6.737541	148.2037	58.85876	3.122439	320.1224	126.6773	53605.92	94.22339	47.24426	11157.55	874.2218	44589.7	12299.93	35953.28
内金	を	支払っ	た	日	は	、	平成	2	1	年	4月	2	7	日	で	あり	、	これら	の
1259.501	2341.285	1529.395	10754573	3105.391	2759.406	515.0094	172.1572	453633.3	180750.4	928913.2	22713.43	333062.8	108684.3	439441.7	7572.135	282.5487	6345.657	165.2635	4777.073
日	から	3	年間	が	経過	し	た	。	したがって	、	被告	ら	は	上記	時効	を	援用	する	。
63.68624	1311.131	499.6885	5871.628	993.1107	90.83181	118167.4	356291.9	378.4269	41.27277	55666.52	179.2601	4.980819	2861.202	1346.384	65.45419	51128.12	38857.43	1.94E+08	67755.79
イ	民事	調停	法	1	9	条	により	、	調停	の	不	成立	による	事件	の	終了	から	2	週間
196.7342	126.0741	26693.09	16142	373398.1	638237	239505.2	54763.53	1291.488	235.4712	6275.138	12330.35	175069.2	37723.72	486.6034	5658.907	2863.772	15724.4	218.3609	1705.38
以内	に	訴え	を	提起	し	なけれ	ば	、	調停	の	申	立て	によって	時効	が	中断	し	ない	と
1327.93	48780.11	287.1027	79686.67	7899.108	276692.4	91384.35	897513.6	2098.209	126.3804	3635.933	5529.626	8215911	67619.91	707.7069	98049.72	7881.757	202047.9	326793.2	14857.27
いう	べき	で	ある	。	本件	事故	は	平成	2	0	年	12月	4	日	に	発生	し	、	原告
3822.546	37576.92	8931.022	223293.5	878720.1	47.86374	66.5146	2400.126	2237.528	201027	240894.1	8889203	8079.858	529630.7	242774.8	7542.753	110.1636	220370.9	884907.4	201.4302
は	平成	2	3	年	11月	2	9	日	に	本件	調停	手続	を	申し立て	、	本件	調停	手続	は
57.78522	387.3364	528417.8	86223.03	8335541	3294.276	988516.3	1827312	9269279	48706.37	130.6586	87.25619	3669.747	8604.11	2286.316	119680.9	93.48465	64.75064	5450.846	13117.13
平成	2	5	年	6月	6	日	に	終了	し	、	原告	は	、	その後	1	ヶ月	以内	の	同年
259.9879	74665.44	97273.89	7031472	18415.14	279948.4	131384.9	16200.08	123.9461	181680	224509	243.9053	51.56936	587.5416	99.52641	2717.393	383.5541	2845.571	19688.51	249.2831
7月	5	日	に	本件	訴訟	を	提起	し	た	の	で	ある	から	、	時効	は	中断	し	、
20614.43	81656.77	21664.27	3175.665	144.7822	59.73817	175024.1	7346.142	75340.27	127058.3	1183.893	15577.97	75223.87	53685.74	15757.18	179.4783	6957.253	10422.96	223847	196031.1
完成	し	て	い	ない	。	遅延	損害	金	原告	は	、	被告	ら	の	求め	に	応じ	て	診断
246.3135	16673.35	135036	8684.118	189725.4	2134.331	30.51668	994791.3	27820.27	2618.206	59.09874	2550.191	175.9593	8.792632	6080.696	290.1922	6432.631	9721.03	9935.856	51.86504
																TP	A	(Confidential word)	
																FP			
																FN		nothing	

Fig.7.7 An example of predicting the result in current precedents after anonymized

Table 7.4 Results of the experiment in various type of anonymized precedents

item	total word	confidential word			normal word			recall	Precision	F1
		apear(TP+FN)	no hit(FN)	hit(TP)	apear(TN)	hit(FP)	no hit			
Rental contact	7800	52	29	23	7748	623	7125	44%	4%	7%
Land cont r act	9600	1588	751	837	8012	806	7206	53%	51%	52%
Traffic accident	2600	67	10	57	2533	280	2253	85%	17%	28%
Traffic accident	8400	100	35	65	8300	831	7469	65%	7%	13%
Rental contact	4000	76	11	65	3924	250	3674	86%	21%	33%
Injury case	12800	543	169	374	12257	1624	10633	69%	19%	29%
Land cont r act	1800	34	26	8	1766	90	1676	24%	8%	12%
Investment receivables	17600	777	177	600	16823	1960	14863	77%	23%	36%
Employment contract	11600	152	31	121	11448	1045	10403	80%	10%	18%
Information disclosure	1600	30	7	23	1570	164	1406	77%	12%	21%
Stock claims	14200	529	82	447	13671	1439	12232	84%	24%	37%
Moving trouble	5200	79	58	21	5121	715	4406	27%	3%	5%
Building surrender	1600	4	0	4	1596	88	1508	100%	4%	8%
Facility admission fee	5800	2	0	2	5798	360	5438	100%	1%	1%
Contract receivables	3800	31	15	16	3769	328	3441	52%	5%	9%
Road maintenance guarantee	8600	34	8	26	8566	516	8050	76%	5%	9%

Therefore, if the recall rate is ≥80%, we consider that a commercialization possibility exists. Our final goal is to automatically replace the anonymous (confidential) words in the actual precedents with the capital letter “A” using our proposed model.

Therefore, we used the test data as the original precedent in which confidential words were not converted to “A.” We experimented with this model to examine whether the model was practical.

7.8.2 An experiment of the proposed model using non-anonymized precedents

We applied the proposed model to 100 cases of plain precedents to analyze the correctness of anonymization and compared the results of "before anonymization" with those "after anonymization by a legal expert," as shown in Table 7.5. Considering this result, after anonymizing, a recall rate of ≥70% was found in 67% of the total,

While before anonymizing it was 33%. Therefore, the recall score was poor "before anonymization" However, precision and F1 scores were better "before anonymization" than "after anonymization". It is necessary to solve the problem explained in the next section for practical use.

Table 7.5. Application of the model to precedents before and after anonymization

Item	Total words	Before anonymizing			After anonymizing		
		Recall	Precision	F1	Recall	Precision	F1
Request land surrender	3299	28%	9%	14%	32%	4%	7%
Rental conteract	2200	37%	19%	25%	33%	6%	10%
Claim for dameges	4600	80%	31%	45%	82%	27%	40%
Advisory contract	7200	58%	5%	10%	50%	1%	2%
Building rental contract	1246	94%	33%	48%	100%	29%	45%
Money lending	2534	58%	44%	50%	25%	22%	24%
Delinquent tax payments	7148	41%	15%	22%	29%	1%	3%
Discipline of lawyer	1044	0%	0%	0%	100%	6%	12%
Internet contract	11280	83%	62%	71%	87%	36%	51%
Illigal residennce(1)	14158	29%	6%	10%	57%	7%	13%
Illigal residennce(2)	3701	58%	11%	18%	74%	10%	17%
Building pollution	4219	55%	9%	16%	25%	2%	3%
Internet sales	1021	72%	12%	21%	92%	24%	38%
Money consumption loan	1007	76%	35%	47%	89%	12%	21%
Aum Shinrikyo	10153	57%	10%	17%	75%	9%	16%

7.8.3 The problem of the proposed model using non-anonymized precedents

In the case of the "Request of land surrender ", the recall was the result that the recall was 28%

and the precision was 9%. The recall score was too bad because many address words have appeared in the precedents. The defect of our new model is that it couldn’t predict the confidential words like an address continuing the long figure. In the case of the "Rental contract", the result was that the recall was 37%, and the precision was 19%.

. One reason why the recall score was low was our proposed model couldn't predict the name of address as the confidential word correctly especially too long address often appeared in the precedents.

Another reason was that "MeCab" sometimes couldn't extract a person's name correctly which appeared by full name. In the case of "Internet contract", the recall score was 83% and the precision score was 62% that showed our model was effective for practical use.

In this case, any person's name appeared in the precedent, our proposed model could predict them correctly.

Considering these results, our model was not enough for practical use at present. However, if many confidential words appeared in the precedents, our proposed model was able to predict the confidential words correctly because of same phrases appear in the Japanese precedents relatively. Furthermore, if “MeCab” can accurately extract the personal name and another proper noun, our proposed model would prevent false detection of the confidential words that were not true. The problem of our model is that couldn't predict the long number address such as 103-123–24-55. If we replace the long number to like "A" as the confidential word in preprocessing, it will be effective for predicting the confidential word. Therefore, it would be able to increase the possibility of practical use.

Fig.7.7 shows an example that our proposed model has anonymized the confidential word for the original non-anonymized precedents.

From Fig.7.7, we can recognize the following results.

① Our model could predict the person’s name, especially the second name as the confidential word correctly. If a person’s name was divided into the first and second names, it couldn’t predict as the confidential word of both names. The accuracy depends on the extracting ability of “MeCab”.

② The word after the punctuation mark is easily recognized as the confidential word whether it is true or not.

た	絵画	六	点	は	、	いずれ	も	児玉	が	昭和	四	四
71530.8818	202.913554	2795.67575	276.034017	867.549931	350.167421	66.7429043	2755.61409	54.5496798	351.810008	399.259305	1478.8218	19096.1101
なる	。	以上	の	とおり	、	児玉	の	総	所得	金額	は	、
1790.79827	17095.9962	49.9824565	639.474614	649.875977	5392.61822	56.4194217	229.122209	591.702055	165.535529	56178.406	4681.21429	301.21201
二	九	〇	〇	万	円	2	雑	所得	の	金額	一	一
59612.64	281461.34	3871002.29	459468.67	351938.378	62012.3735	301.853528	1371.58579	614.391753	3127.00965	7466.02447	218.753367	816.57306
〇	万	円	イ	野村証券	から	受領	し	た	謝礼	五	〇	〇
402.887123	339.8193	24473.8572	303.510602	56.7377255	38.9246098	220.390675	3228.6991	284381.474	205.908507	9633.32955	9872.05049	1411.8291
二	〇	六	八	万	二	五	〇	〇	円	ジャパン	ライン	から
36781.462	467848.521	461336.567	18034.4259	227902.218	7256.40814	223252.86	1615387.12	65971.1753	282400.872	336.963701	163.311883	235.091412
万	円	ウ	曽根	啓介	から	受領	し	た	謝礼	五	〇	〇
862.546287	30150.0392	318.154351	75.2715667	3025.09886	173.737913	295.815756	24587.0543	544469.424	303.107272	10502.298	12474.7515	3533.54527
ある	。	ア	東海興業	から	受領	し	た	謝礼	二	〇	〇	〇
20189.0985	158754.023	11.2138536	61.6432074	366.294141	166.249702	3499.04767	235037.972	442.770411	7502.2475	15962.3082	8690.52305	1338.67977

TP FP TN

Fig.7.7 An example of practical use

7.9 The problem of the proposed model applying to the plain precedents (non-anonymized) and the solution for practical use

7.9.1 The problem of the proposed model using non-anonymized precedents

Our proposed model was unable to correctly predict the names of addresses as confidential words, especially as addresses in the precedents were often very long. This problem is one reason why the recall score was low. Another reason was that MeCab sometimes could not correctly extract a person’s name when the full name was present. From these results, we can conclude that our model is not currently sufficient for practical use. If many confidential words appeared in the precedents, the proposed model correctly predicts them.

7.9.2 A solution to problems in the experimental results for practical use

Confidential words were sometimes replaced with non-English letters, such as the Greek letters γ or β, in which case our proposed model was unable to predict them as confidential words (see Fig.7.8). Besides, addresses in the precedents appearing as “xx-xx-xx” were not predicted as confidential words. When these letters appeared in the precedents, they would be properly recognized as confidential words if the preprocessing algorithm could be improved. As shown in Table 7.6, when this improved algorithm was applied to the precedents that contained many addresses and Greek letters, the recall score improved by about 4%–49%, compared to that before

Table 7.6. The result of preprocessing improvement

ID	Item	Before improvement			After improvement		
		recall	Precision	F1	recall	Precision	F1
3344	Request land surrender	32%	4%	7%	42%	3%	6%
3556	Building pollution	25%	2%	3%	74%	6%	12%
3337	Aum Shinrikyo	75%	9%	16%	79%	11%	20%

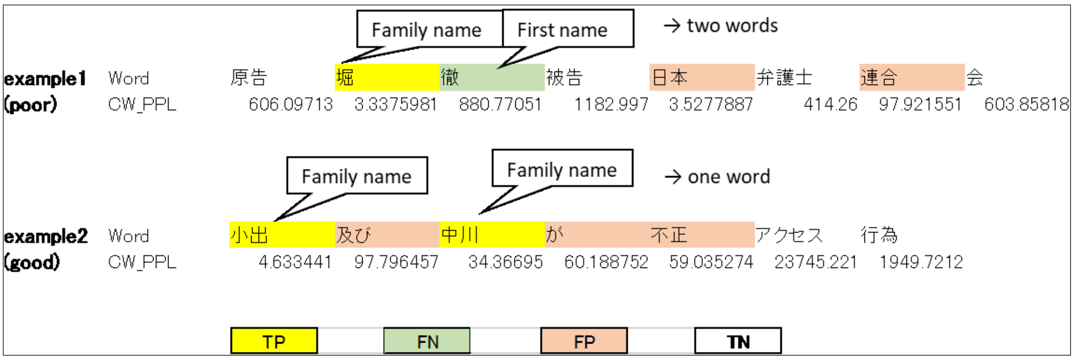


Fig.7.8. Example of predicting names of people

the improvement. If we improve MeCab's ability to accurately extract the names of people, the recall score will increase. In particular, MeCab was unable to extract first and family names as one word; hence, our model was unable to recognize names as the confidential word. We have shown an example to explain this in Fig. 7.8.

In Fig. 7.8, Example 1 shows that the proposed model was able to predict the family name but not the first name because the name was extracted as two words. In contrast, Example 2 shows that it was correctly able to predict two names because each name was extracted as one word.

Another issue was that it was unable to recognize the long numbered address as confidential words. If MeCab can extract the address as a proper noun, recognition probability will improve (see Fig. 7.9), or the preprocessing algorithm will be improved, the model can recognize them as confidential words.

Another example that the Greek letters such as " δ " are contained in the precedent is shown in Fig.7.10. In this case, our model couldn't be recognized " δ " as a confidential word.

入谷	九	丁目	2	7	番地	4	,	2	7	番地	3
574.3243	57.97388	105.7458	88.31534	2997.067	2656.133	20.5286	728.7321	39.76389	290.2797	28.78283	94.88714

Fig.7.9 Example of predicting the long numbered address

杉並	区	δ	×	丁	目	×	×	番地
573.9007	1967.426	223.3996	169.4301	1260.279	9311.73	42.18243	2031.932	3124.297

Fig.7.10 Example of predicting the Greek letter

Improving the preprocessing algorithm, we could get a good recall score as shown in table 7.6.

Next, we applied the improved preprocessing algorithm to many actual precedents some of which include address. We showed the experimental result in Table 7.7.

As a result, we got a recall rate was over 70% in about 90% of tested precedents, compared with before improvement of preprocessing that showed it was over 70% in 67% of tested precedents

Table 7.7 Experimental result of the improved preprocessing applied to many actual precedents

No.	item	total word	confidential word			normal word			recall	Precision	F1
			Apear (TP+FN)	no hit (FN)	Hit (TP)	Apear (TN+FP)	Hit (FP)	no hit (TN)			
3294	Bill wages	1083	19	1	18	1064	96	968	95%	16%	27%
5950	Request for restoration	11151	349	59	290	10802	1301	9501	83%	18%	30%
2653	Yield the building	1250	42	10	32	1208	104	1104	76%	24%	36%
2649	Movable property delivery request	1081	51	6	45	1030	168	862	88%	21%	34%
3340	Information disclosure request	10595	1008	86	922	9587	1621	7966	91%	36%	52%
3337	Yield the building（Aleph）	10836	112	17	95	10724	1375	9349	85%	6%	12%
5313	Lawyer discipline	957	9	2	7	950	62	888	78%	10%	18%
3311	Claim for damages	4613	405	81	324	4208	690	3518	80%	32%	46%
3556	Request for cancellation of administrative disposition	4239	12	3	9	4227	291	3936	75%	3%	6%
3307	Appeal for compensation claim	7065	9	2	7	7056	713	6343	78%	1%	2%
5316	Debt seizure	7169	10	4	6	7159	477	6682	60%	1%	2%
3338	Yield the building	1887	22	3	19	1865	173	1692	86%	10%	18%
5952	Claim for damages	5192	51	14	37	5141	706	4435	73%	5%	9%
5953	Claim for damages	1755	3	0	3	1752	138	1614	100%	2%	4%
3344	Request for delivery of building	3305	31	10	21	3274	412	2862	68%	5%	9%
3349	Claim for reimbursement	2744	104	4	100	2640	264	2376	96%	27%	43%
2511	Mandatory withdrawal of decision	3763	28	5	23	3735	214	3521	82%	10%	17%
3551	Request for cancellation of deportation order	13903	133	31	102	13770	1454	12316	77%	7%	12%
5951	Collection claim appeal	2589	32	19	13	2557	215	2342	41%	6%	10%
5950	Request for restoration	4561	245	34	211	4316	699	3617	86%	23%	37%
3548	Claim for damages	965	14	1	13	951	55	896	93%	19%	32%
3552	Request for cancellation of building confirmation	2539	58	8	50	2481	219	2262	86%	19%	31%
3859	Request for cancellation of provisional release disapproval disposition	7008	171	46	125	6837	758	6079	73%	14%	24%

Furthermore, if MeCab can accurately extract names and proper nouns, our proposed model will prevent the false detection of confidential words and improve the precision and F1 score. This will increase the potential for practical use. These issues are summarized in Table 7.8.

Table 7.8. Issues of our model

	issues	measure
1	Long numbed address can't be recognized correctly	Upgrade the "MeCab" and "preprocessing"
2	Person's name sometimes can't be recognized correctly	Upgrade the "MeCab" and "preprocessing"
3	Precision and F1 score are low	Upgrade the “MeCab” and Further study

7.9.3 The experimental results of the proposed model applying to more plain precedents after improving the preprocessing algorithm

Because the effect of improving the recall score was obtained by improving the preprocessing algorithm, we conducted further experiments that we applied our proposed model to more plain precedents.

At first, we have applied our model to more various types of plain precedents after anonymizing manually. As a result, a recall score of ≥70% was found in 86% of the total which was 67% of the total before preprocessing improvement. That means our model could predict confidential words with high accuracy. The experimental results are shown in Table 7.9.

Only in collection claim appeal case (No.5951; the color-coded area in brown), recall score was 41%. It's so bad. The model sometimes couldn't predict 'A' that was a person's name as confidential words but miss-predicted the next word of 'A' as the confidential word. We couldn't figure out the cause yet. But we suppose it's a rare case.

Next, we have applied our model to the same plain precedents but before anonymization to confirm the possibility of commercialization.

As a result, the recall score of ≥70% was found in 50% of the total which was 33% of the total before preprocessing improvement as shown in table.7.9. The recall score improved by 20% before the preprocessing improvement. However, there were some issues described below.

In the debt seizure case (No.5316), the recall score was very low because the model couldn't predict many 'account number' as confidential words. These account number was replaced to '○○○○' when they were anonymized manually. So, the model couldn't learn them as confidential words. If they were replaced with 'A', our model could predict them as confidential words (Fig.7.11).

In the wage billing case (No.3294), the recall score was high for 'after anonymization' but low for 'before anonymization'. The person name '芝博史' appeared many times in this original precedent (non-anonymized) adding the prefix '亡' before the person name '芝博史'. However,' 亡' means 'dead' not 'person name' (Fig.7.12). The words '亡芝博史' replaced with '亡 A', by anonymization work manually. So, the model could predict them as a confidential word with high accuracy after anonymization. But when +'亡芝博史' appeared in the precedents before anonymization, the model couldn't predict '芝博史' as the personal as 'Unknown'. Therefore, the recall score was low in the non-anonymized precedent.　In this case, if MeCab would extract these pattern correctly, the recall score would be better.

割引	国債	0	0	0	6	3	9	-	0
437.64765	3634.1218	1218.1548	391.49883	326.80758	176.50509	599.13893	491.36032	1348.0261	41.794789
							FN	TP	

Fig.7.11 Example of predicting the long account number

2	<UNK>	博	史	は	.
57.4778	114.762	298.63	19.6047	14.79	908.235

Fig.7.12 Example of predicting the person name adding prefix

Precision was less than 30% in each preceding. F1 score was 2%-62%, and it was below 50% in most cases. It means the model often miss-predict the non-confidential words as the confidential words. However, our purpose of this study is that the model can predict the confidential words without omission. So, the recall score should be prioritized than precision score. Then we evaluate the experimental result using the F2 score. The equation is as follows.

$$F2 = \frac{5(precision)(recall)}{4precision+recall} \qquad (7\text{-}15)$$

F2 which emphasis the recall may be useful in evaluating the model instead of F1 but F2 score is also low because precision score was very low. It is necessary to increase the precision score to put the model to practical use. The measures to improve the precision score may be to upgrade the MeCab to extract the confidential word as proper noun accurately or to adopt the new model like BERT. These are further study.

Table 7.9. Application of the model to the plain precedents after preprocess improvement

No.	item	total word	After anonymization				Before anonymization			
			Recall	Precision	F1	F2	Recall	Precision	F1	F2
3294	Bill wages	1083	95%	16%	27%	47%	39%	14%	21%	29%
5950	Request for restoration	11151	83%	18%	30%	49%	85%	23%	37%	56%
2653	Yield the building	1250	76%	24%	36%	53%	64%	25%	36%	49%
2649	Movable property delivery request	1081	88%	21%	34%	54%	90%	21%	35%	55%
3340	Information disclosure request	10595	91%	36%	52%	70%	90%	47%	62%	76%
3337	Yield the building (Aleph)	10836	85%	6%	12%	25%	72%	7%	13%	25%
5313	Lawyer discipline	957	78%	10%	18%	33%	83%	10%	17%	33%
3311	Claim for damages	4613	80%	32%	46%	62%	81%	31%	45%	62%
3556	Request for cancellation of administrative disposition	4239	75%	3%	6%	13%	46%	2%	4%	9%
3307	Appeal for compensation claim	7065	78%	1%	2%	5%	64%	1%	3%	6%
5316	Debt seizure	7169	60%	1%	2%	6%	8%	1%	2%	4%
3338	Yield the building	1887	86%	10%	18%	34%	38%	6%	11%	19%
5952	Claim for damages	5192	73%	5%	9%	20%	61%	9%	16%	29%
5953	Claim for damages	1755	100%	2%	4%	10%	71%	4%	7%	15%
3344	Request for delivery of building	3305	68%	5%	9%	19%	38%	4%	7%	14%
3349	Claim for reimbursement	2744	96%	27%	43%	64%	84%	34%	48%	65%
2511	Mandatory withdrawal of decision	3763	82%	10%	17%	33%	85%	17%	28%	47%
3551	Request for cancellation of deportation order	13903	77%	7%	12%	24%	57%	7%	12%	22%
5951	Collection claim appeal	2589	41%	6%	10%	18%	49%	9%	15%	26%
3548	Claim for damages	965	93%	19%	32%	52%	82%	25%	38%	56%
3552	Request for cancellation of building confirmation	2539	86%	19%	31%	50%	44%	14%	22%	31%
3859	Request for cancellation of provisional release disapproval disposition	7008	73%	14%	24%	40%	71%	28%	41%	55%

Chapter 8

Future tasks

8.1 Introduction

As we could the good accuracy for predicting the confidential word using our new proposed model "the Bi-directional LSTM LR using POS tag", I tried to apply the new proposed model to various precedents. However, there are still challenges to apply to actual precedents. The following research is required for practical use.

8.1.1 Improvement of the preprocessing

① It is necessary to improve the preprocessing algorithm for long address numbers and telephone numbers that can be represented by continuous numbers. If address numbers and telephone numbers are detected and replaced to 'A' in the preprocessing, the recall sore will be further improved.

② If account numbers in the precedents are detected by judging from the context,

③ Upgrade the MeCab or other morphological analyzer

The proposed model sometimes couldn't predict the confidential words when MeCab sometimes couldn't extract a person's name correctly when the full name appeared because the name was extracted as two or three words. If the MeCab has upgrade the extraction ability of person name, the proposed model will be great step for practical use. Or it may be necessary to adopt other morphological analyzer.

④ Exclude the stop words such as particles and auxiliary verbs

Stop words are words that are excluded from processing because they are common and useless in natural language processing. For example, functional words such as particles and auxiliary verbs ("わ", "の", "です", "ます", etc.)

8.1.2 Discriminating the legal terms which has often appeared in the precedents

The legal terminology such as “plaintiff” or “defendant” are often miss-predicted as confidential words and the precision score are worth in our experiment. If these words will be excluded in the preprocessing, the accuracy will be further improved.

8.1.3 Adopting the new evaluation method instead of perplexity

The low perplexity (PPL) model is a good prediction model. The evaluation by PPL shows the number of choice of target word that connect to words that occur at some point with equal probability. It does not include the criteria for the word's own error. Some words are easy to make mistakes and some are not. Generally, shorter words are easier to make mistakes. As the PPL does not include criteria such as word length or ease of interaction with other words a decrease in PPL does not always lead to an immediate improvement in recognition accuracy. So the new evaluation method may be needed.

8.1.4 A study of new method

We have been studying the prediction method of confidential word using neural network so far. Considering the hardware, I think we need an instrument that it has a large amount of memory and can calculate at high speed. On the other hand, considering the software, I will pay attention to the state-of-art BERT (Bidirectional Encoder Representations from Transformers) technology which has developed by Google [8-1]. BERT is a natural language model that are learning from before and after the context. There are two steps using BERT: pre-training and fine-tuning. During pre-training, the model is trained on unlabeled data over different pre-training tasks. For fine-tuning, the BERT model is first initialized with the pre-trained parameters, and all of the parameters are fine-tuned using labeled data from the downstream tasks. Each downstream task has separate fine-tuned models, even though they are initialized with the same pre-trained parameters.

BERT has a great feature that can understand the context of sentences, that is, grammar. This is what we are looking for and I think it is suitable for detecting the confidential word in Japanese precedents.

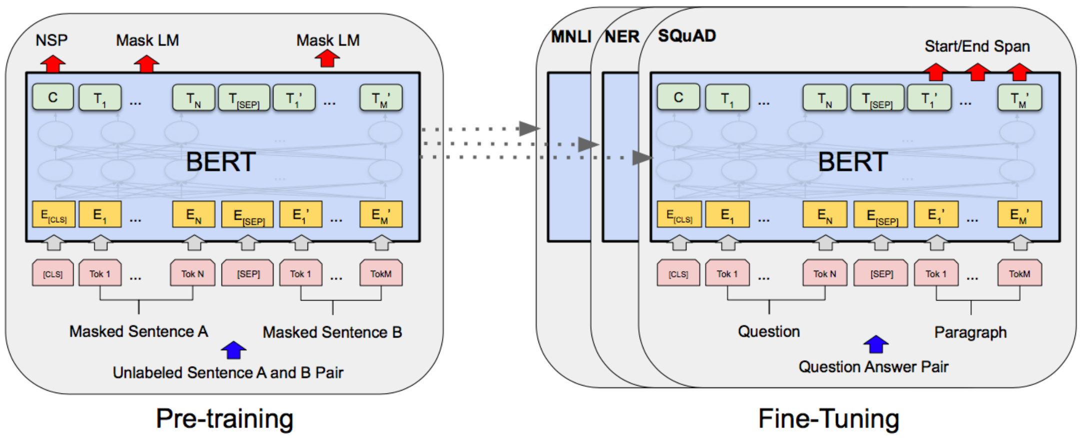


Figure 1: Overall pre-training and fine-tuning procedures for BERT. Apart from output layers, the same architectures are used in both pre-training and fine-tuning. The same pre-trained model parameters are used to initialize models for different down-stream tasks. During fine-tuning, all parameters are fine-tuned. `[CLS]` is a special symbol added in front of every input example, and `[SEP]` is a special separator token (e.g. separating questions/answers).

Fig.8-1 Overall pre-training and fine-tuning

Chapter 9

Conclusion

Introducing IT (Information Technology) into the Japanese judicial system is already described in future investment strategies 2017. However, specific introduction has not been advanced yet. Because Japanese society has various issues that anyone can be participating in the judicial field. Especially, privacy is one of the most critical issues for communication. If we can automatically hide confidential words, the communication becomes safer, and we can realize one piece of CogInfoCom. For that purpose, we developed a technology to predict confidential words to target Japanese judicial precedents.

However, it is not easy to completely anonymize confidential words such as personal names and locations. Named entity recognition (NER) is probably the first step for information extraction to seeks to locate and classify named entities in the Japanese precedents into predefined categories such as the names of persons, locations. Nevertheless, the problem is that the cost of manually making a corpus is high.

One technique to protect privacy is to find confidential words in a file or on a website and convert them into meaningless words. In related works, some named entity recognition (NER) technique to extract the anonymous word has been published so far [9-1] [9-2] [9-3]. However, they need to prepare the proper noun dictionary that is updated to the latest issue. It takes a high cost.

The other method used in machine learning can learn the pattern of a named entity by preparing the corpus. There are HMM (Hidden Markov Model), and CRF (Conditional random fields) for a machine language.

However, HMM is only dependent on every state and its corresponding observed object. HMMs-based method is still limited because it is difficult to model arbitrary, dependent features of the input word sequence. Also, CRF is highly computationally complex at the training stage of the algorithm. It makes it very difficult to re-train the model when newer data becomes available.

On the other hand, the neural network is useful to predict the confidential word that is intelligent enough to anonymize confidential words automatically.

Based on Japanese judicial precedents, we have already proposed a recognition technique for confidential words using a neural network.

At first, we proposed the Bi-directional LSTM-LR model that was effective for detecting the target words in long sequential words. But the accuracy of detecting the confidential words (CW_PPL) in the precedents was worse. To improve the CW_PPL score, we attempted to add another information to the neural network that was Part of Speech (POS) tag. We considered the CW_PPL score would be well by entering the two types of information, that were words and POS tag into the neural network and learning. Then we proposed a new model. It was the Bi-directional LSTM-LR combined the Part of speech (POS) tag extracted by "MeCab" that is a kind of CRF (Conditional Random Field) and a Japanese morphological analyzer.

Further, we reviewed the algorithm of preprocessing. We found that many confidential words, replaced to a wide width alphabet had been missed in the previous algorithm. So, we improved the algorithm. Then, we experimented again using the model. As a result, the CW_PPL score was much improved. We got an accuracy improvement of 88% for detecting the confidential word (CW_PPL) and a 21 % improvement for detecting the target word (PPL) compared to the previously proposed model (Bi-directional LSTM-LR). Then, we evaluated our proposed model with the evaluation index, such as "recall" or "precision," to find out if our proposed model deserved practical use.

As a result, we confirmed our proposed model could predict the confidential word for practical use in some precedents that the recall score showed from 69% to 84%. After anonymizing the recall score of ≥70% was found in 67% of the total, while before anonymizing it was 33%. However, the overall results were not satisfactory. So, we analyze the cause and improving the preprocessing algorithm, experimented again using the more actual precedents. Finally, we got the recall score of ≥70% was in approximately 90% of tested precedents. It means a better recall score compared to that before improved preprocessing algorithm.

Then, we experimented with the non-anonymized precedents to examine for practical use. As a result, the recall score of ≥70% was found in 50% in total. This score is better than before improving the preprocessing algorithm. Our model proved to be practical in some cases. Totally, our model was not enough for practical use at present. However, as it is proved our new proposed model using the neural network combined POS tag is effective for the predicting the confidential words in Japanese precedents, further improvements to this model would be worth it. We would make further improvement for the preprocessing algorithm and upgrading the "MeCab", etc. If we would get the better experimental results to predict the confidential words from now on. We will establish an automatic detector tool for the confidential words.

I have been researching a method for predicting the confidential words in the Japanese precedents in order to develop a Cyber court in Japan that is far behind in judicial field for ICT conversion compared with foreign countries. Although this research may be a small step toward the adoption of ICT in Japanese trials, I feel this study will be effective for Japanese Cyber court in future.

In the news of Yomiuri Shimbun on Feb.4, 2020, it was reported that a web conference was introduced in nine courts nationwide in Japan to conduct civil trial procedures on the Internet. Even in civil trials in Japan, which have been delayed compared to overseas, the use of IT will accelerate in the future.

Acknowledgements

I would like to express the deepest appreciation to prof. Atsushi Ito for his consistently support. I owe my deepest gratitude to him. Without his persistent help and comments, I could not have been continued this study as a working student.

I would also like to appreciate to Dr. Toyama for kindly advice and making the Tex-style document.

I would like to express my gratitude to Mr. Kiryu, KDDI Co. for giving me a great deal of corporation in programming work.

I would also like to express my gratitude to Managing Director Yamasawa and Mr. Ishikawa, TKC Co. showing me a legal data such as precedents and giving me useful comments.

I am deeply grateful to Prof. Kasahara, Toin Yokohama University for giving me insightful comments and suggestions.

I am also grateful to Dr. Hatano and members of his Lab for making the paper in Tex style.

And I would like to say deeply thanks to Dr. Watanabe for continuously warm encouragement.

I am particularly grateful for the assistance given by members of Ito Lab and Toyama Lab.

Finally, I would like to extend my deepest gratitude to my late wife for giving me courage during her lifetime.

Appendix1

The algorithm of the Bi-directional LSTM combined POS tag

Input data

Input corpus ：Japanese precedents

Preprocessing

Input data

$w_1^j = w_1 w_2, \cdots\cdots w_j$ (no space)

*data cleaning # delete punctuation mark and unnecessary words

*insert space between word (w_1)and word (w_2)

*extract the POS tag of the word

Output data

$w_1^j = w_1 , w_2, \cdots\cdots, w_j$ (with space) ：*word*

$p_1^j = p_1 ,p_2,\ \cdots\cdots,\ p_j$ ：*POS*

*p_1 is a POS corresponding to w_1

Main process

Input data

/parameter

hidden_layer_size = 100

batch_size = 200

chunk_size = 10

epochs = 100

learning_rate = 0.001

forget_bias = 1.0

/Word

Backward data $(\mathrm{w_i})_{\mathrm{i}}^{bj}$: $w_1, w_2, \cdots w_{10}$

Forward data $(w_i)_i^{fj}$: $w_{-1}, w_{-2}, \cdots w_{-10}$

/POS

Backward POS data $(p)_i^{bj}$: $p_1, p_2, \cdots p_{10}$

Forward POS data $(p_i)_i^{fj}$: $p_{-1}, p_{-2}, \cdots p_{-10}$

#chunk size (window size) =10

Step1→ Initialize LSTM

Step2→ get the embedding vector for input data

for j in range (chunk size)

$\boldsymbol{X_b} = (w_b)_{i=1}^{j}$: *Word(backward)*

$\boldsymbol{X_f} = (w_f)_{i=1}^{j}$: *Word(forward)*

$\boldsymbol{Y_b} = (p_b)_{i=1}^{j}$: *POS(backward)*

$\boldsymbol{Y} = (p_f)_{i=1}^{j}$: *POS(forward)*

Step3→ Ready lstm cell (tensorflow)

bw_ lstm = BasicLSTMCell ($\boldsymbol{X_b}$) : word (backward)

fw_ lstm = BasicLSTMCell ($\boldsymbol{X_f}$) : word(forward)

p_bw_ lstm = BasicLSTMCell ($\boldsymbol{Y_b}$) : POS (backward)

p_fw_ lstm = BasicLSTMCell ($\boldsymbol{Y_f}$) : POS (forward)

: backward outputs $h_t^{(b)} = Wh_t + Wh_{t+1} + b$

: forward outputs $h_t^{(f)} = Wh_t + Wh_{t-1} + b$

: POS backward outputs $s_t^{(b)} = Ws_t + Ws_{t+1} + b$

: POS forward outputs $s_t^{(f)} = Ws_t + Ws_{t-1} + b$

* W: weight, b: bias

Output

#the last output is the model's output

outputs=concat(bw_outputs,af_outputs,H_bw_outputs,H_fw_outputs)

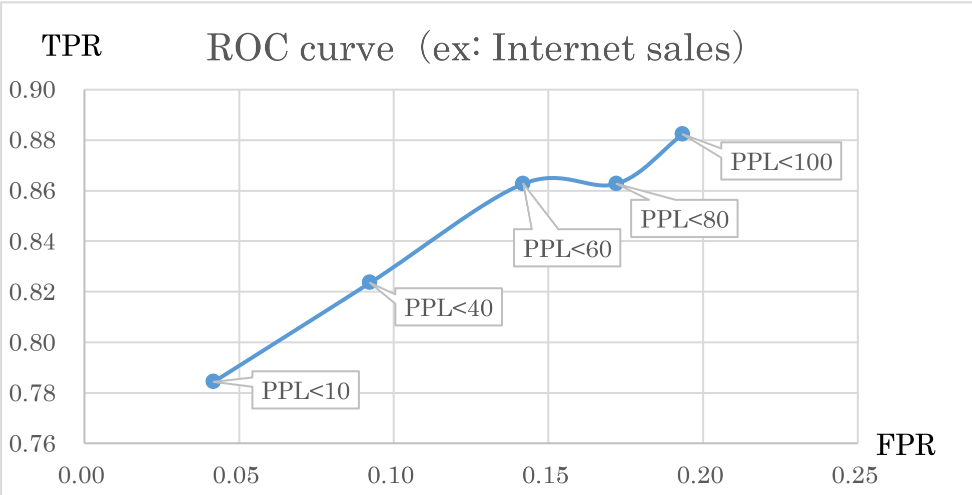
outputs $\boldsymbol{Y} = (h_t^{(b)} + h_t^{(f)} + s_t^{(b)} + s_t^{(f)})$

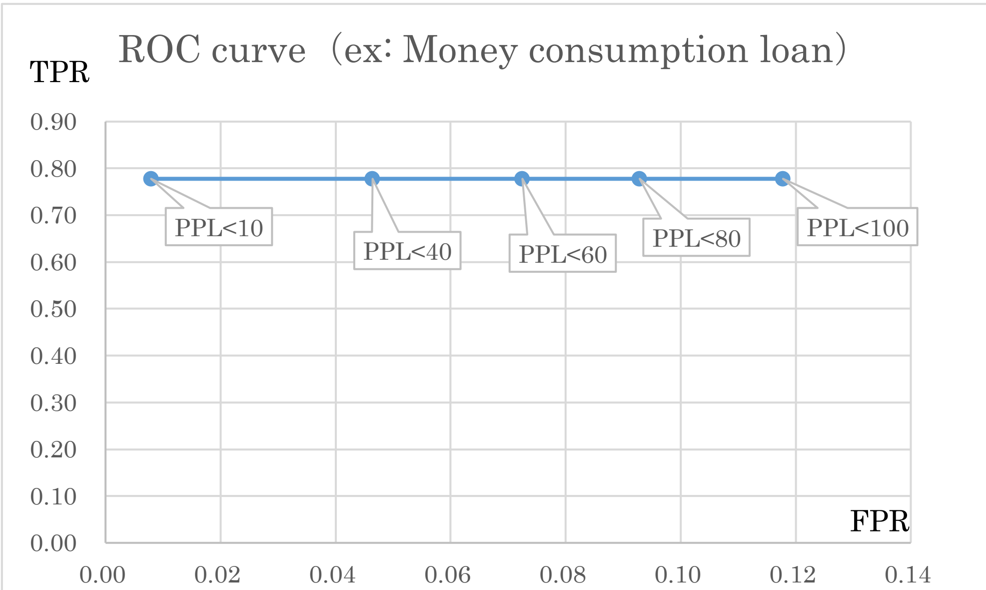
output= Y[outputs] $\times W + b$ ：loss

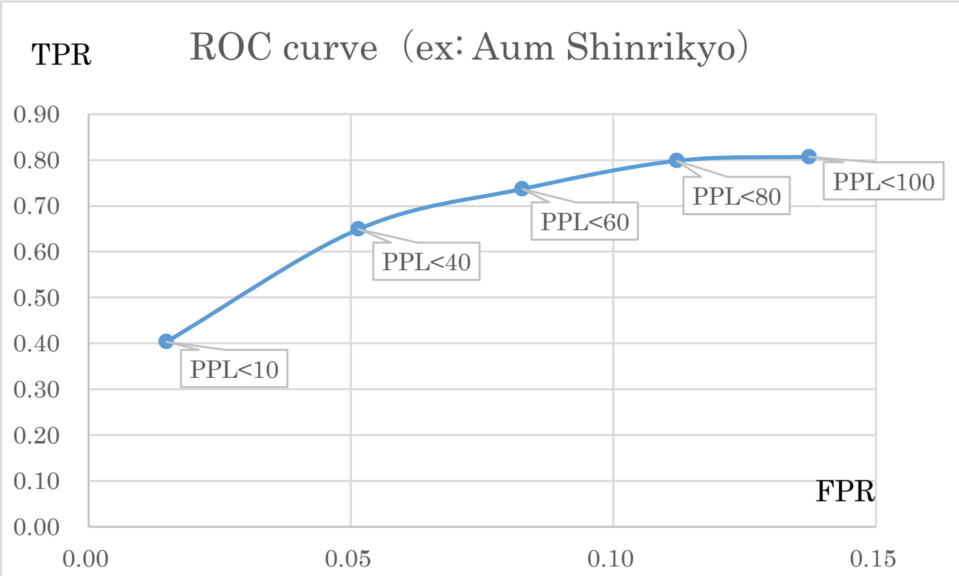
$$Y = (y_i)_i^j = (y_1, y_2, \cdots\cdots\cdots, y_i) \quad : word$$

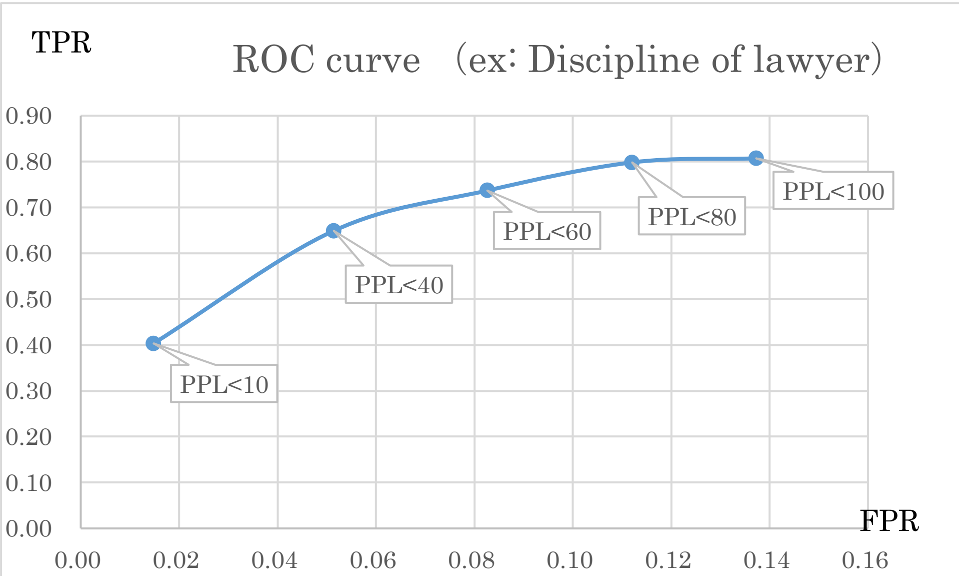
Appendix 2

ROC curve for each actual precedents






TPR
ROC curve（ex: Aum Shinrikyo）
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
0.00
0.05
0.10
0.15
FPR
PPL<10
PPL<40
PPL<60
PPL<80
PPL<100


TPR
ROC curve （ex: Discipline of lawyer）
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
0.00
0.02
0.04
0.06
0.08
0.10
0.12
0.14
0.16
FPR
PPL<10
PPL<40
PPL<60
PPL<80
PPL<100

Appendix3

Future investment strategy 2018

Improving the system and environment for promoting digital government

1 . Promotion of IT in court procedures

Court procedures in civil litigation while respecting the judiciary's autonomous judgment. The following efforts will be carried out in stages.

・ First, under the current law, the Judicial Office will hold a web conference, etc. from next year.56 Started trial and operation of dispute arrangements etc. to be used in extreme cases, We hope that we will work to improve issues such as arranging issues.

・ Next, oral legislation date, etc., which will make necessary legislative arrangements and do not require the appearance of related parties and aim to start a new system from around FY2021. The Ministry of Justice will promptly consider the Legislative Council during the next fiscal year consider and prepare in advance. The Judiciary has a prompt expecting action, the executive branch will take necessary measures.

・ Furthermore, necessary environment development such as legal maintenance and system construction will be carried out.

The Ministry of Justice has decided to implement the necessary legal for realization, promptly consider the consultation of the Legislative Council during the next fiscal year.

・ Also, the Ministry of Justice has set a schedule for realizing online

After considering the Judicial Office's examination and efforts for improving the environment, we will consider this next year.

Bibliography

[1-1] Ministry of Internal Affairs and Communications Japan, "2019 white paper information and communications in Japan", Chapter3- 2019

[1-2] http://law.wm.edu/academics/intellectuallife/researchcenters/clct/ [Sep, 5, 2017]

[1-3] https://www.supremecourt.gov.sg/default.aspx?pgID=361#10 [Sep, 5, 2017]

[1-4] JURISIN, Japan, Oct 2013

[1-5] JURISIN, Japan, Oct.2014

[1-6]https://www.theverge.com/tech/2017/8/18/16167836/china-cyber-court-hangzhou-internet-disputes

[1-7] Takehiko Kasahara, "Cybercourt - Court Technology", Journal of the Japanese Society for Artificial Intelligence, vol. 23, No. 4, pp 513-, Jul. 2008]

[1-8] Takehiko Kasahara, "Court Technology for Civil Procedure", Festschrift for the seventieth birthday of Prof. Takeshi Kojima II, pp. 961-, [Sep. 2009]

[1-9] "Can you judge?", NHK Special, [Sep. 2005.]

http://www6.nhk.or.jp/special/detail/index.html?aid=20050213 [Sep, 5, 2017]

[1-10] http://www5.cao.go.jp/keizai-shimon/kaigi/minutes/2017/0609/shiryo_07.pdf, [Sep, 5, 2017]

[1-11] http://www5.cao.go.jp/keizai-shimon/kaigi/minutes/2017/0609/shiryo_07.pdf, [Sep, 5, 2017]

[1-12] https://www.theverge.com/tech/2017/8/18/16167836/china-cyber-court-hangzhou-internet-disputes

[1-13] Masaki Takase, et al. Named Entity Recognition for Animation-Related Words Using CRF. The Language Processing Society 20th Annual Conference [March 2014]

[1-14] LAI, Siwei, et al. How to generate a good word embedding. IEEE Intelligent Systems, 2016, 31.6: 5-14.

[2-1] https://www.shojihomu.or.jp/kenkyuu/saiban-it [Dec. 2019]

[2-2] Prime minister of Japan and His cabinet, Future investment strategy 2018, p112

[2-3] Study group on IT conversion in court procedures-Japan Economic Revitalization Headquarters,
http://www.kantei.go.jp › singi › keizaisaisei › saiban [Dec.2017]

[2-4] Information Technology in Judiciary (Electronic Litigation)

[3-1] H.Kawashima, et.al "Comparative Legal Basic Study on Creative System Construction of "e-Support Court", "
http://www.taf.or.jp/files/items/file

[3-2] Sugimura, IT in court procedures in Singapore and America [Dec.2017]
https://www.kantei.go.jp/jp/singi/keizaisaisei/saiban/dai2/siryou5.pdf

[3-3] Kasahara, IT in court procedure in Europe [Dec.2017]
https://www.kantei.go.jp/jp/singi/keizaisaisei/saiban/dai2/siryou6.pdf

[3-4] Hiraoka, IT in court procedure in Korea [Dec.2017]

https://www.kantei.go.jp/jp/singi/keizaisaisei/saiban/dai2/siryou4.pdf

[3-5] Onodera, "IT and internationalization of intellectual property and trial in China" [July.2019]
http://www.icclc.or.jp/icclc-news/news_62_1.pdf

[3-6]Ministry of Justice; http://www.moj.go.jp/content/001322234.pdf

[3-7] Electronic communication between lawyers and courts:
https://www.infogreffe.com/informations-et-dossiers-entreprises/actualites/communication-par-voie-electronique-entre-avocats-et-juridiction-premiers-deploiements-rpva-tc-i-greffes.html

[5-1] Sendai Bar Association, "Opinion on civil IT procedures"
https://senben.org/wp-content/uploads/2019/07/IT_ikensho.pdf

[5-2] Atsushi Hiraoka, "Is Japan a law-abiding country?" [July, 2010]
http://senkensoi.net/column/2010/07/2774 [July,2010]

[6-1] Baranyi, P., Csapo, A. (2012). Definition and Synergies of Cognitive Infocommunications, Acta Polytechnica Hungarica, Vol. 9 No. 1, pp.67-83

[6-2] Baranyi, P., Csapo, A., Sallai, G. (2015). Cognitive Infocommunications (CogInfoCom), Springer International, ISBN 978-3-319-19607-7.

[6-3] Lai, S., et al. How to generate a good word embedding. IEEE Intelligent Systems, 2016, 31.6: 5-14.

[6-4] http://www.courts.go.jp/app/hanrei_jp/search1 [Dec, 31, 2017]

[6-5] Bengio, Y., Ducharme, R., Vincent, P., & Jauvin, C. (2003). A neural probabilistic language model. Journal of machine learning research, 3(Feb), 1137-1155.

[6-6] Mikolov, T., et al. Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781, 2013.

[6-7] Masaki Takase, et al. "Named Entity Recognition for Animation-Related Words Using CRF", The Language Processing Society 20th Annual Conference, March. 2014

[6-8] Bengio, Y., Ducharme, R., Vincent, P., & Jauvin, C. (2003). A neural probabilistic language model. Journal of machine learning research, 3(Feb), 1137-1155.

[6-9] Mikolov, T., et al. Recurrent neural network based language model. InterSpeech 2010 pp1045-1048 [Sep.2010]

[6-10] Bojanowski, P., et al. Enriching word vectors with subword information. arXiv pre-print arXiv:1607.04606, 2016.

[6-11] Shunya Ariga; Yoshimasa Tsuruola. Extension of synonym words according to context by vector representation of words. The 21st Annual Conference of Natural Language Proceedings (NLP2015), 2015, 752-755

[6-12] http://www.tkc.jp/law/lawlibrary [Dec, 31, 2017]

[6-13] http://taku910.github.io/mecab/ [Mar, 9, 2018]

[6-14] LAI, Siwei, et al. How to generate a good word embedding. IEEE Intelligent Systems, 2016, 31.6: 5-14.

[6-15] Yuya Kiryu, Atsushi Ito and Masakazu Kanazawa, Recognition Technique of Confidential Words Using Neural Networks in Cognitive Infocommunications. ActaPolytechnicaHungarica [Sep9,2018]

[6-16] http://kaldi-asr.org/doc/dnn3_scripts_context.html

[6-17] LSTM-Based Hierarchical Denoising Network for Android Malware Detection (Security and Communication Networks Vol. 2018, Article ID 5249190, 18 pages)

[6-18] Masakazu KANAZAWA, Atsushi ITO, Yuya KIRYU, et al.'Improve accuracy of predicting confidential words in judicial precedents':Technical Committee on Thought and Language (TL) [July,28,2018]

[6-19] http://taku910.github.io/mecab/ [Mar, 9, 2018]

[6-20] John Lafferty, Andrew McCallum, & Fernando Pereira. (2001). Conditional Random Fields : Probabilistic Models for Segmenting and Labeling Sequence Data. In Proceedings of the 18th International Conference on Machine Learning 2001 (pp. 282– 289).

[7-1] MeCab: Yet Another Part-of-Speech and Morphological Analyzer. http://mecab.sourceforge.net/

[7-2] 奥村　学監修、高村大也著「自然言語処理のための機械学習入門」コロナ社

[7-3] Masakazu Kanazawa, Atsushi Ito, Kazuyuki Yamasawa, Takehiko Kasahara, Yuya Kiryu and Fubito Toyama; 'Method to Predict Confidential Words in Japanese Judicial Precedents Using Neural Networks With Part-of-Speech Tags'　Infocommunications Journal 2020.1st issue, DOI: 10.36244/ICJ.2020.1.3 Volume XII、No.1, pp12-16

[7-4] Masakazu Kanazawa, Atsushi Ito, Kazuyuki, Takehiko Kasahara, uya Kiryu;' Predicting the confidential words in Japanese judicial precedents using the neural network, 2019 International Conference on Computational Science and Computational Intelligence (CSCI): pp.337-342: DOI 10.1109/CSCI49370.2019.00066

[7-5] https://towardsdatascience.com/understanding-auc-roc-curve-68b2303cc9c5

[8-1] Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina 'Toutanova BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding arXiv:1810.04805

[9-1] Li, J., Sun, A., Han, J., Li, C.: A Survey on Deep Learning for Named Entity Recognition. cs.CL (2018).

[9-2] Misawa, S., Taniguchi, M., Miura, Y., Ohkuma, T.: Character-based Bidirectional-LSTM CRF with words and characters for Japanese Named Entity Recognition Proceedings of the First Workshop on subword and Character Level Models in NLP. Fuji Xerox Co. Ltd. pp. 97102 (2017).

[9-3] Ken Y, Koru I, Shoko W, Eiji, W : The 31st Annual Conference of the Japanese Society for Artificial Intelligence.2017. DOI: https://doi.org/10.11517/pjsai.JSAI2017.0_2J2OS16a4

List of Publications

Journal papers

- Recognition Technique of Confidential Words Using Neural Networks in Cognitive Infocommunications

  Yuya Kiryu, Atsushi Ito, Masakazu Kanazawa

  Acta Polytechnica Hungarica, Vol. 16, Issue 2, pp.129-143 (2019).

- Method to Predict Confidential Words in Japanese Judicial Precedents Using Neural Networks with Part-of-Speech Tags

  Masakazu Kanazawa, Atsushi Ito, Kazuyuki Yamasawa, Takehiko Kasahara, Yuya Kiryu, Fubito Toyama

  Infocommunications Journal Vol. XII, No. 1, pp.17-25. (2020).

International Conference Papers

- A study on improving accuracy to predict confidential words in judicial precedents using the neural network

  Masakazu Kanazawa, Atsushi Ito, Kazuyuki Yamasawa, Takehiko Kasahara

  Twelfth International Workshop on Juris-informatics (JURISIN 2018), 13 pages, Kanagawa, Japan, 2018.11.

- Predicting the confidential words in Japanese judicial precedents using the neural network

  Masakazu Kanazawa, Atsushi Ito, Kazuyuki Yamasawa, Takehiko Kasahara, Yuya Kiryu

  2019 International Conference on Computational Science and Computational Intelligence (CSCI), pp.337-342, Las Vegas, USA, 2019.12

- A study on a trial to anonymize plain Japanese precedents by using Machine Learning Technology

  Masakazu Kanazawa, Atsushi Ito, Kazuyuki Yamasawa, Takehiko Kasahara, Yuya Kiryu, Fubito Toyama

  Fourteenth International Workshop on Juris-informatics (JURISIN 2020), 14 pages, Online workshop, 2020.11

Oral treatise

- Improve accuracy of predicting confidential words in judicial precedents

  <u>Masakazu Kanazawa</u>, Atsushi Ito, Yuya Kiryu, Kazuyuki Yamasawa, Takehiko Kasahara

  Institute of Electronics, Information and Communication Engineers Technical Committee on Thought and Language [July,28,2018]