




Masked prediction and interdependence network of the law using data from large-scale Japanese court judgments

Ryoma Kondo^{1,2} · Takahiro Yoshida² · Ryohei Hisano^{2,3} 

Accepted: 26 September 2022 / Published online: 21 October 2022
© The Author(s) 2022

Abstract

Court judgments contain valuable information on how statutory laws and past court precedents are interpreted and how the interdependence structure among them evolves in the courtroom. Data-mining the evolving structure of such customs and norms that reflect myriad social values from a large-scale court judgment corpus is an essential task from both the academic and industrial perspectives. In this paper, using data from approximately 110,000 court judgments from Japan spanning the period 1998–2018 from the district to the supreme court level, we propose two tasks that grasp such a structure from court judgments and highlight the strengths and weaknesses of major machine learning models. One is a prediction task based on masked language modeling that connects textual information to legal codes and past court precedents. Another is a dynamic link prediction task where we predict the hidden interdependence structure in the law. We make quantitative and qualitative comparisons among major machine learning models to obtain insights for future developments.

Keywords Legal code prediction · Court judgments · Text mining · Network mining

Ryoma Kondo, Takahiro Yoshida and Ryohei Hisano have equally contributed to this work.

✉ Ryohei Hisano
hisanor@g.ecc.u-tokyo.ac.jp

Ryoma Kondo
kondo@mlab.t.u-tokyo.ac.jp

Takahiro Yoshida
t_yoshida_74@outlook.com

¹ Graduate School of Engineering, The University of Tokyo, 7-3-1, Hongo Bunkyo-ku, Tokyo 113-0033, Japan

² The Canon Institute for Global Studies, 5-1, Marunouchi 1-chome, Chiyoda-ku, Tokyo 100-6511, Japan

³ Graduate School of Information Science and Technology, The University of Tokyo, 7-3-1, Hongo, Bunkyo-ku, Tokyo 113-0033, Japan

1 Introduction

Court judgments (or court opinions) not only are a by-product of litigation processes resolving real-world disputes, but also contain crucial information on the interpretation and interdependence of the law (including both legal codes and past court precedents) in the current legal system. This holds in countries adopting both common and civil law systems. For instance, in the US, the famous quote “At the heart of the First Amendment is the recognition of the fundamental importance of the free flow of ideas and opinions on matters of public interest and concern.” Hustler magazine (1988) highlights one supreme court interpretation of what the First Amendment is all about. The same kind of quote is abundant in the Japanese court judgments as well. For instance, the Supreme Court Grand Bench precedent, “Whether or not a penal provision contravenes Article 31 of the Constitution because of ambiguity should depend on whether a person with ordinary judgment can understand the criterion, by which he can decide whether the provision applies to an act in a specific case” (Grand bench of the supreme court of 1973 1975) sets up the golden standard of when Article 31 of the Japanese constitution can be invoked. This court precedent itself is heavily cited by other court judgments to decide whether the legal code applied to a real-world case brought in court is against Article 31. This fact also implies that Article 31 of the Constitution is often used in conjunction with the interpretation given by Grand bench of the supreme court of 1973 1975 highlighting one standard interdependence structure in the legal system.

For the legal profession, understanding the connection between textual information written in court judgments and the law is essential from both the judiciary and legislative perspectives. From the judiciary side, knowing precisely what the legal code means and how actual cases are handled in litigation is essential for lawyers, prosecutors, and judges. Lawyers and prosecutors can polish their court strategy, while judges can review past court precedents to deliver final judgments. This explains why there are professional legal search engine services and many professional books summarizing important court precedents for professional use. From the legislative side, even in civil law systems such as Japan, the legal code is not the entire story. First, in Japan, the supreme court has the right to review and provide a conclusive interpretation of the law’s constitutionality. Second, past court precedents (or “*hanrei*” in Japanese) offer a necessary interpretation of the statutory laws and how they should be used in the litigation process. Even though the Diet is defined as the sole law-making institution in Japan in the Constitution, this interpretation and supplementation step of the court could be seen as having a “law-making” (“*kihan-teiritsu*” in Japanese) role in influencing future court judgments as well as future legislation. Moreover, the importance of past court precedents to cultivate the practices of the court in countries with common law systems such as the US is apparent (i.e., case law). Thus, to understand fully how the law is interpreted in society, we need quantitative technologies to data-mine how laws interact with each other and evolve in the courtroom.

In Japan, tens of thousands of court judgments are written every year. Seeing this corpus of court judgments as text data creates a unique opportunity

to data-mine both the connection between textual information written in court judgments and the statutory laws and the interdependence (i.e., interaction network) between them. The former task could be achieved by defining a classification problem connecting textual information to statutory laws and past court precedents. As court judgments are written formally, this could be performed by creating a training data set from the corpus of past court judgments using natural language processing techniques. The latter is also quite simple to implement, especially for Japanese court judgments. Usually, there are, on average, 10–20 law articles and past court precedents used to deliver the final decision in each case. By creating a clique graph from co-occurrence patterns, we can aggregate these cliques to develop an interaction network that summarizes the interdependence structure of the law. Moreover, the dynamic link prediction problem defined in this network could be seen as completing the missing interactions in the law (e.g., an interaction among the law that is plausible but not yet written because no real-world cases that would invoke the interaction have been brought into the courtroom), making it exciting as both a technical and practical legal task.

Hence, in the present paper, we propose two novel tasks to achieve the goals mentioned above by using a unique data set that includes approximately 110,000 court judgments from the district to the supreme court level spanning the period 1998–2018, provided by a firm that offers professional legal search engine services in Japan. The first task is a masked legal language prediction task that predicts legal code (including article numbers) and past court precedents from masked sentences that hide them.

Some previous works have focused on similar sentences with citation information. For instance, Zhang and Koppaka (2007) focused on sentences that include citations of past court opinions in the US and coined them as the “reason for citation.” They also attempted to locate the sentence that caused the citing court opinion to cite the particular court opinion and coined the cited sentence as the “text of interest.” Using these two key sentences, they proposed a graph traversal approach for search engine purposes. Another line of works (Sadeghian et al. 2018; Shulayeva et al. 2017) classified the reason for citation itself. For instance, the reason for citation could be classified into several types, such as legal basis, authority, definition, and exception (we refer to Sadeghian et al. 2018 for the complete list). Although these works are interesting, the difference with our work is that our focus is on the prediction of statutory laws and past court precedents’ names.

Many previous works have focused on prediction tasks in the legal domain (Liu et al. 2015; Sulea et al. 2017; Wang et al. 2018; Nguyen et al. 2018; Medvedeva et al. 2020; Dadgostari et al. 2020; Tagarelli and Simeri 2021). The Competition on Legal Information Extraction/Entailment (COLIEE) is a popular workshop held at the International Conference on Artificial Intelligence and Law (ICAIL). The tasks in this workshop include legal case retrieval tasks using Canada case law data sets and statutory law retrieval tasks using legal bar exam data sets from Japan¹. Many excellent papers have been written using these data sets, providing researchers with

¹ <https://sites.ualberta.ca/rabelo/COLIEE2021/>.

opportunities to conduct legal data science research (Morimoto et al. 2017; Nanda et al. 2017; Yoshioka et al. 2021; Nguyen et al. 2021). Another similar but different legal code prediction task is the work of Dadgostari et al. (2020). By viewing legal searches as a prediction problem, Dadgostari et al. (2020) proposes a learning-based model that (1) predicts citations in US court opinions from its semantic content and (2) predicts the search results generated by human users. In the same vein, Tagarelli and Simeri (2021) uses deep learning technologies to predict the most relevant Italian Civil Code of a query sentence. Legal judgment prediction has also attracted many researchers. Recently, using decisions from the EU Court of Human Rights as a test case, Medvedeva et al. (2020) showed that 75 percent accuracy could be achieved in predicting the violation of nine articles of the EU Convention on Human Rights. The use of deep learning methods has also entered the area of legal predictive tasks (Yamakoshi et al. 2019; Chalkidis and Kampas 2019). Several pretrained models have also been proposed in the legal domain (Chalkidis et al. 2020; Tagarelli and Simeri 2021). The difference from this research is that we focus on predicting both statutory laws and past court precedents' names using masked sentence data from a large corpus of real-world court judgments in Japan.

The second task we explore in this paper is a dynamic link prediction task using a network created by aggregating clique graphs from the co-occurrence patterns in each court judgment. Network analysis of the law is not a new concept. For instance, the work of Fowler and Jeon (2008) focuses on the citation network of past court opinions in the US, and Coupette et al. (2021) focuses on cross-reference networks in the legal code using data sets from both the US and Germany. Coupette et al. (2021) shows that the complexity of the legal code has been increasing over the past 30 years. They also provide a profile analysis of the law using both textual information and a reference network structure. Sakhaee and Wilson (2021) performed a similar reference network analysis on the New Zealand legal code. By contrast, Mazzega et al. (2009); Boulet et al. (2011, 2018) focused on the French case and La Cava et al. (2021) focused on the Italian case. Network analysis is also used in the field of quantitative comparative law. By analyzing reference networks for several European countries, Badawi and Dari-Mattiacci (2019) showed that the structure of reference networks tends to be similar among countries with similar influences. Koniaris et al. (2018) proposed the "Legislation Network" approach to quantify the interdependence structure using EU legal sources for legislation purposes. Furthermore, there is now even a patent granted in the US that offers legal path analysis for a policy disruption early warning system (Alex et al. 2020) based on the technologies proposed in Lyte et al. (2015).

The contributions of the present paper can be summarized as follows:

- To the best of our knowledge, this is the first study to data-mine large-scale court judgment documents (approx. 110,000) from the district to the supreme court level in Japan.
- We propose a novel legal masked language prediction task to connect textual information (reason for citation) to legal codes and past court precedents.
- We give both an extensive quantitative and qualitative analysis of major machine learning models and show that deep learning technology models lead to highly

predictable outcomes. We also illustrate limitations and possible directions for future research.

- Using the co-occurrence patterns in a court judgment, we propose a novel dynamic link prediction task that identifies the possible set of interactions within the law. This task is essential both as a technical exercise and a practical legal task.
- We show that using a simple network model already leads to good performance. Still, a model that uses textual and network information leads to the best predictive performance.
- We also provide a qualitative assessment of the learned embedding from the best performing model.

2 Japanese legal system and data set

2.1 Japanese legal system

Before describing our data set, we briefly explain the Japanese legal system to provide context to the unfamiliar reader. Japan adopts a civil law system (similar to Germany and France). Statutory laws come first in the civil law system and are complemented by court precedents. Compared with the common law counterpart (the US and the UK), the Japanese legal system relies heavily on the precise interpretation of what is specified in the legal code. The role of the juridical system is to give its interpretation. The Japanese legal code consists of the Constitution, statutory laws, government ordinances, and ministerial ordinances. The Constitution is the most powerful, and any statutory laws or ordinances that violate the Constitution are invalidated. Statutory laws come next, in which the Diet is the only institution responsible for creating and making changes. The administrative branch enacts government and ministerial ordinances to complement the details of statutory laws for regulatory purposes.

The supreme court has the final say on the interpretation of the law and the Constitution. Moreover, the supreme court sometimes provides an interpretation of a specific legal code, not only to solve the case but also to avoid future confusion. Thus, supreme court judgments that define the interpretation of the law (including the Constitution) are regarded as essential court precedents. These important court precedents act as if it is part of the legal code (this law-making aspect is called “*kihan-teiritsu*” in Japanese). Hence, it is essential to consider significant court precedents to understand the legal code in its entirety.

Japan adopts a three-trial system, with the supreme court on top, followed by high courts and district courts. In general, the district courts first consider cases brought to the juridical system. If the case is appealed, it will be heard by the high court, and if the case is appealed again, the supreme court gives the final judgment. While the district and high courts focus on the facts and application of the law, the supreme court mainly focuses on evaluating the law’s interpretation and constitutionality. Many cases reaching the supreme court level are rejected based on the lack of need to consider the law’s interpretation and constitutionality.

Court judgments typically consist of two parts. The “*shubun*” part comes first, stating the decisions made by the court. It typically consists of several sentences that mainly indicate whether the plaintiff’s claim is upheld in civil cases and whether the accused is guilty in criminal cases, without getting into the details of the court’s logic leading to a particular decision. The “*shubun*” part is followed by the reasoning part, which describes the facts and legal reasoning behind the decisions. As we are interested in analyzing the logic of a court judgment, here, we focus only on the reasoning part, ignoring the “*shubun*” part and all the meta-information (e.g., dates, case number, judges name) written in court judgments².

2.2 Summary of the data set

In this paper, we use a comprehensive data set on the court judgments in Japan spanning 20 years, from 1998 to 2018. The data were provided by a legal search engine company in Japan (TKC Corporation). In the provided data set, nouns corresponding to personal information, such as personal and company names, were replaced with pseudonyms. In Japan, privacy protection is prioritized over information disclosure, making only a handful of court judgments publicly available. This situation is one of the reasons why a large-scale data analysis of Japanese court judgments has not been performed. From the reasoning section of the court judgments, we extracted all the sentences containing citations of the statutory law at the article number level and past court precedents using regular expression techniques. For the statutory law names, we created a list of all the legal codes for the statutory laws and used it to determine whether a particular phrase is indeed a statutory law³. As in the eyecite code for the free law project⁴, there are key lexical rules to cite past court precedents in Japan as well. We created expression patterns based on those rules and extracted the court precedents’ names using regular expression techniques. Our approach to focusing on citations of statutory laws and court precedents is similar to Zhang and Koppaka (2007), in which such sentences as the “reason for citation” were coined. As there were a lot of spelling variants (especially court precedents), we also performed rule-based normalization of the law and court precedents’ names, an example of which is shown in Table 1.

We provide a summary of the extracted “reason for citation” sentences in Table 2. We restricted our summary to law articles and court precedents that appeared more than four times in our analysis. To illustrate the difference in the number of articles and unique court precedents among the civil and criminal court cases, we counted the number of sentences for each category (statutory law articles appearing in civil cases, civil court precedents, statutory law articles appearing in criminal cases, and criminal court precedents). Sentences denote the number of reasons for citation

² For more details of the Japanese legal system, see (Jones 2020).

³ There are cases where some legal codes have changed their meaning over time. Taking these variations would almost likely increase the model’s performance, but since the amount of work needed to construct a table of all changed legal codes for the studied period is massive, we leave this for future work.

⁴ <https://github.com/freelawproject/eyecite>.

Table 1 Citation extraction and normalization

Original text	Extracted citation	Normalized citation
<p>これらの規定に違反しないことは当裁判所の判例(最高裁昭和22年(れ)第119号同23年3月12日大法院判決・刑集2巻3号191頁)とすることであるから。</p>	<p>最高裁昭和22年(れ)第119号同23年3月12日大法院判決・刑集2巻3号191頁</p>	<p>最高裁判所昭和23年3月12日判決</p>
<p>Past court precedents of the Supreme court has determined that these provisions are not violated (Supreme Court, Grand Bench judgment of March 12, 1948, 1947 (Re) No. 119, Keishu Vol.2, No.3, at 191).</p>	<p>Supreme Court, Grand Bench judgment of March 12, 1948, 1947 (Re) No. 119, Keishu Vol.2, No.3, at 191</p>	<p>Supreme Court, Judgment of March 12, 1948</p>

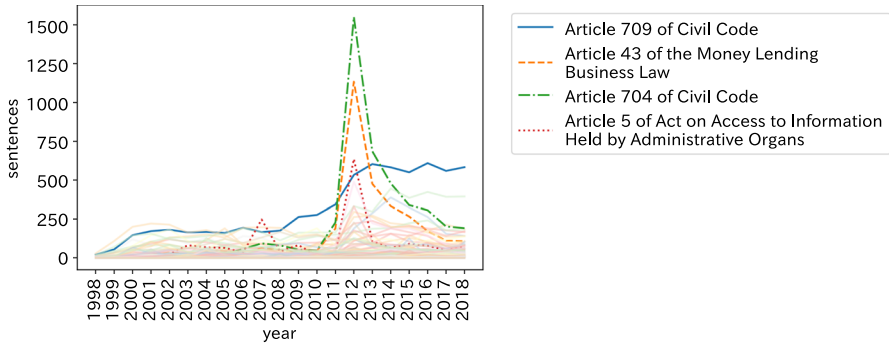


Fig. 1 Number of citations of civil articles

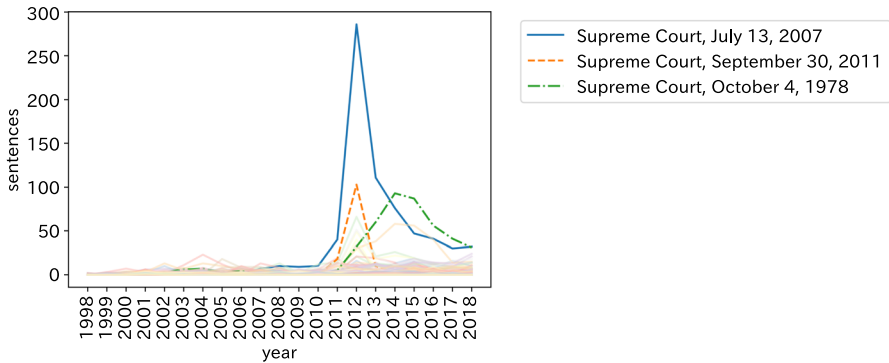


Fig. 2 Number of citations of civil precedents

Table 2 Summary of the data set

Category	Sentences	Labels
Civil articles	600,558	8797
Criminal articles	34,845	630
Civil court precedents	47,369	1561
Criminal court precedents	1009	76

sentences, while labels denote the number of unique law articles or court precedents found in our data set. We see that civil court cases have the highest number of sentences for both sentences and labels, while criminal court precedents have the lowest number of sentences.

The number of sentences citing each label follows a very heavy- and long-tailed distribution (Figs. 3, 4, 5, 6). For example, the number of sentences citing Article

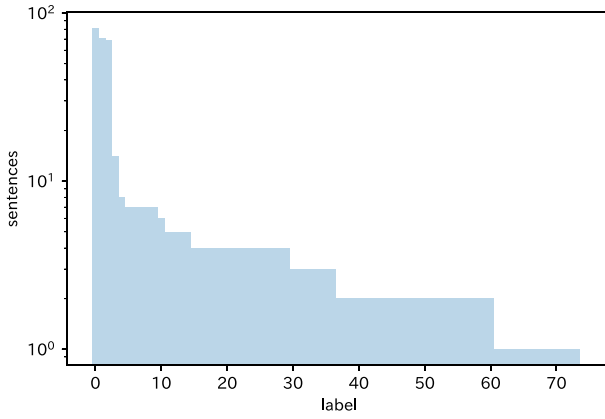


Fig. 3 Distribution of data from criminal court cases

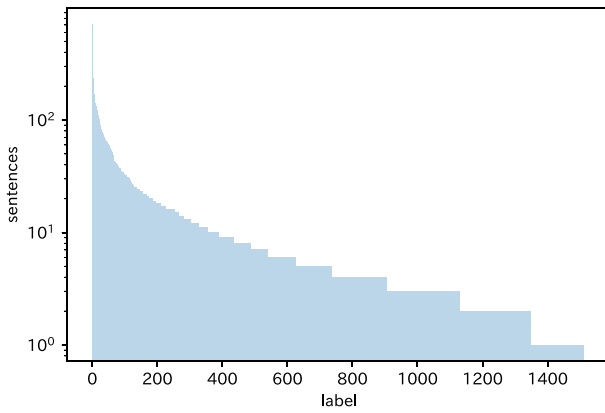


Fig. 4 Distribution of data from civil court cases

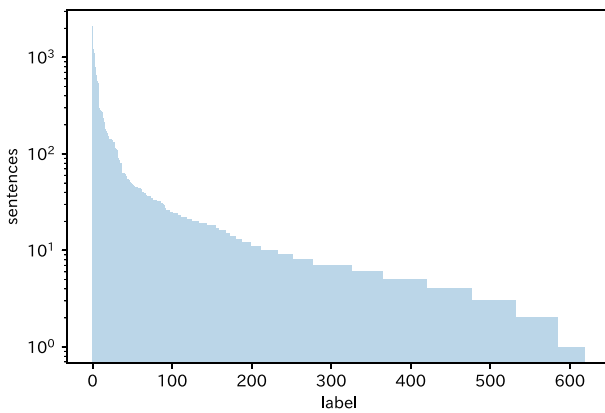


Fig. 5 Distribution of data from criminal articles

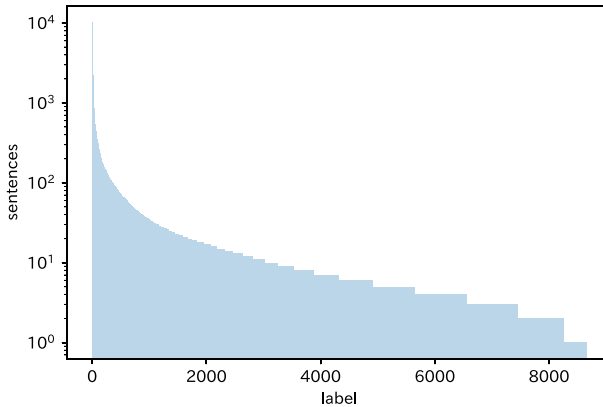


Fig. 6 Distribution of data from civil articles

709 of the Civil Code is 21,380⁵. By contrast, many classes appear only five times. Moreover, the frequency of each class varies over time. Figures 1 and 2 show a time series plot summarizing the frequency of statutory laws found in civil cases and civil court precedents. We can see that statutory laws and court precedents related to overpayment refund claims skyrocketed around 2012.

Overpayment refund claim cases were a massive game changer for the legal professions and consumer finance companies in Japan. In the past, two different statutory laws set the upper limit of the consumer finance interest rate. One was the Interest Rate Restriction Act (20 percent), and the other was the Investment Act (29.2 percent). This inconsistency created the so-called “gray zone interest rate” where consumer finance companies could choose which upper limit to obey. As a profit-maximizing entity, the consumer finance companies obviously chose the higher one, and this was further reinforced by Article 43 of the former Money Lending Business Act⁶, which stated that an interest rate over 20 percent and below 29.2 percent would be legitimate as long as the borrower agreed to pay the higher interest rate “voluntarily” (任意に (*nin-i-ni*)). However, on Jan 13, 2006 (Judgment of the second petty bench of 2004 2006), the supreme court gave an interpretation of this phrase “voluntarily” (任意に (*nin-i-ni*)) in the Money Lending Business Act that was much stricter than what the consumer finance companies had expected. Furthermore, on July 13, 2007 (Judgment of the second petty bench of 2005 2007), the supreme court gave another critical court precedent obliging the consumer finance companies to be virtually forced to repay the borrower with an additional legal interest rate of 5 percent a year, defined in the former Article 404 of the Civil Code. These decisions rendered all the interests that the consumer finance company had charged in the past illegitimate, and the borrower obtained the right to recover any excess payments they had

⁵ Article 709 of the Civil Code defines the requirement for compensation for damage and thus is often used.

⁶ This should not be confused with the current Article 43, which is entirely different.

paid in the past plus the legal interest rate. As many of the following court cases that cited the overpayment court precedents were civil lawsuits that reclaimed the excessive payments as much as possible, court precedent July 13, 2007, was cited more than the more basic court precedent of June 13, 2006, as we could confirm from Fig. 2.

As this example shows, the number of citations for the statutory laws and court precedents also varies with time, sometimes changing drastically because of important court precedents given by the supreme court (and obviously by the changes in the legal code). Moreover, some statutory laws and court precedents cope with a similar issue but differ significantly in terms of what interpretation is given. Therefore, for machine-assisted search engines to be beneficial in real-world applications, it is crucial to predict the exact statutory laws or court precedent's name from texts.

3 Masked language prediction

In this section, we focus on predicting the statutory laws and court precedents' names from the "reason for citation" masking them. Although this is a simple problem, it tests whether we can find a meaningful connection between textual information written in court judgments and the law. We compare several machine learning methods, from a simple model using simple features to a more advanced model using complex semantic features and advanced machine learning techniques. We provide both quantitative and qualitative comparisons of the models to highlight their strengths and weaknesses.

All models' primary input is the "reason for the citation" with the citation target replaced by a string implying a mask token. Specifically, we define three types of masks: [MASK_LAWNAME], which masks the statutory laws name, [MASK_ARTICLE], which hides the article number, and [MASK_PRECEDENT], which masks the court precedents. If multiple statutory laws or court precedents exist in a single text, we choose the predictions from the top according to their score.

3.1 Summary of the compared methods

Conceptually, all the tested models can be divided into an encoder and a decoder. The encoder is the part that converts masked sentences into quantitative vectors that capture the meaning of the sentences. The decoder is the part that outputs class labels, which in the current setting are the statutory laws and court precedents. We compare several methods for the encoders to identify the best way to capture the semantic structure of sentences. For most of the models, we used a gradient boosting technique (Chen and Guestrin 2016), which is an ensemble learning method based on decision trees⁷. We also tested the Text-To-Text Transfer Transformer (T5) (Rafael et al. 2019), an end-to-end deep learning model with its own encoder-decoder

⁷ Using random forest showed slightly inferior performance.

structure, to explore whether changing the decoder part affects the predictive accuracy. The methods used in our experiments are described below.

3.1.1 Encoder

The simplest encoder for a sentence is a vector summarizing its **word frequency** (i.e., bag-of-words approach). The drawback of this approach is that the vector does not reflect the word order. We separated Japanese sentences into words using morphological analysis and counted the word frequency for each sentence. We restricted the counting to terms that appeared at least 10 times in the data set.

Multilingual universal sentence encoder (USE) (Yang et al. 2019) is a convolutional neural network model that reflects the order of words. The model has been pretrained using the Standard Natural Language Inference (SNLI) data set, a standard public data set for natural language, and is available from Google to be used without additional local training. The SNLI data set is augmented into 16 languages, including Japanese, using Google Translate. Specialized words that did not appear in the pretraining step are treated as unknown words (i.e., unknown tokens (UNK)) and are ignored. The strength of this approach is that it tries to capture the semantic meaning of a sentence compared with the simpler word frequency approach. The method's weakness is that it is not easy to interpret the meaning of each dimension compared with the word frequency approach. This point will be elaborated on in the Results section.

We also tried several other encoders. **Doc2vec** (Le and Mikolov 2014) is a method for sentence embedding that uses a simple neural network. This method converts sentences into vectors based on word2vec, which converts words into vectors using continuous bag-of-words and skip-gram techniques. The neural network was trained from the initial state using the masked sentences in the training data. **Latent Dirichlet allocation (LDA)** (Blei et al. 2003) is a type of clustering method for extracting topics from documents. A topic is represented as a probability of the distribution of words in a training data set. We first trained LDA using only the data in the training set. The likelihood of each topic for the test set can be obtained by inputting sentences to the trained model, which is used as an embedding vector like the ones above.

3.1.2 Decoder

For all the encoders listed in the previous section, we used gradient boosting (Chen and Guestrin 2016) as the decoder. Gradient boosting is an ensemble method based on a decision tree and has shown state-of-the-art performance in many classification problems. As already mentioned in the encoder section, we also tested using the T5 (Raffel et al. 2019), which is an end-to-end deep learning model with its own encoder and decoder. The T5 model uses an attention mechanism to embed the meaning of the input sentence into a vector. The model is pretrained on a data set called C4, a set of web pages published on the Internet and filtered for pages written in English and containing expletives. We used T5-base, which is a domain adaptation pretrained on the Japanese Wikipedia.

One point worth mentioning about the T5 model is that it does not treat the prediction problem as a classification problem. Instead, given an input sentence, it outputs a word or sentence as in generation tasks. This characteristic makes the loss function different from the gradient boosting counterpart. In the gradient boosting case, if we missed the article number of the statutory law (e.g., predicting Civil Code 709 as Civil Code 710), it would simply obtain a loss of failing to predict the statutory law. However, in the T5 model, if we mistook the article numbers such as the Civil Code 709 (“民法 709”) and the Civil Code 710 (“民法 710”), the model would realize that the prediction was a close match up to the third character (“民法7”). As most codes on the same topic tend to line up together, this advantage makes it easier for the T5 to achieve better performance.

3.2 Experimental settings

Seventy percent of the data set was divided into a training data set, and the rest into a test data set. We use accuracy as one of our evaluation metrics, defined as the ratio of simple correct predictions to the number of ground truth labels. More precisely, let \mathbf{S}_i^{gt} be the set of ground truth labels in the i th input sentence. Using the model’s score for each label given an input \mathbf{X}_i , we denote the set of labels in the upper k of this score as $T(\mathbf{X}_i, \theta, k)$. Using this notation, we define the accuracy of the model θ as follows:

$$Accuracy = \sum_i^N \frac{1}{|\mathbf{S}_i^{gt}|} \sum_{l \in T(\mathbf{X}_i, \theta, |\mathbf{S}_i^{gt}| + m)} C(l, \mathbf{S}_i^{gt}) \quad (1)$$

$$C(l, \mathbf{S}_i^{gt}) = \begin{cases} 1 & (l \in \mathbf{S}_i^{gt}) \\ 0 & (else) \end{cases} \quad (2)$$

where m is an evaluation parameter representing the number of additional predictions besides the number of masked labels. For the T5 model, it was difficult to obtain the likelihood (or similar score) because it does not treat the problem as a natural language generation problem. Hence, for the T5 model, we only report the accuracy when $m = 0$.

Under the above definition of accuracy, class imbalances are not fully considered. For instance, Article 709 of the Civil Code, which appears the most frequently, consists of about 3.6 percent of the [MASK_LAWNAME] in the civil legal code problem (we used 8,797 labels in our setting). To deal with this problem, we use the F-measure previously used in Tagarelli and Simeri (2021) as the second evaluation metric. We define the micro-averaged F-measure as follows. Let precision P_l be the number of correct predictions for the label l . The number of correct answers among the numbers included in the ground truth for the label l is defined as recall R_l .

$$F_\mu = \frac{1}{|L|} \sum_{l \in L} \frac{2P_l R_l}{P_l + R_l} \quad (3)$$

We also define the macro-averaged F-measure as follows:

Table 3 Accuracy scores

Model	Category	Accuracy ($m=0$)	Accuracy ($m=1$)	Accuracy ($m=2$)	Accuracy ($m=3$)	Accuracy ($m=4$)
onehot+XGB	Civil code	57.84	63.56	66.09	67.67	68.75
onehot+XGB	Criminal code	80.06	84.1	86.18	87.6	88.41
onehot+XGB	Civil case	68.12	74.29	77.22	78.86	80.04
onehot+XGB	Criminal case	68.32	72.61	75.25	76.24	77.89
USE+XGB	Civil code	51.98	59.16	62.89	65.38	67.21
USE+XGB	Criminal code	71.84	76.84	79.66	81.35	82.60
USE+XGB	Civil case	61.54	69.05	72.56	74.59	76.08
USE+XGB	Criminal case	65.02	70.63	74.59	76.24	77.89
T5	Civil code	70.32	–	–	–	–
T5	Criminal code	54.13	–	–	–	–
T5	Civil case	24.11	–	–	–	–
T5	Criminal case	8.37	–	–	–	–
LDA+XGB	Civil code	52.24	58.46	61.43	63.27	64.56
LDA+XGB	Criminal code	70.77	75.22	77.89	79.71	80.95
LDA+XGB	Civil case	60.66	66.57	69.40	71.12	80.95
LDA+XGB	Criminal case	36.30	39.27	40.26	41.91	43.89
doc2vec+XGB	Civil code	0.09	0.11	0.18	0.33	0.83
doc2vec+XGB	Criminal code	16.29	19.88	22.99	25.94	28.73
doc2vec+XGB	Civil case	19.03	20.32	21.14	21.91	22.64
doc2vec+XGB	Criminal case	27.06	33.33	35.97	37.95	39.93

$$F_M = \frac{2P_\mu R_\mu}{P_\mu + R_\mu} \quad (4)$$

$$P_\mu = \frac{1}{|L|} \sum_{l \in L} P_l, R_\mu = \frac{1}{|L|} \sum_{l \in L} R_l$$

As shown in the above equation, the F-measure is an evaluation index that increases when it has high prediction accuracy over many classes. We use these evaluation metrics in the following section.

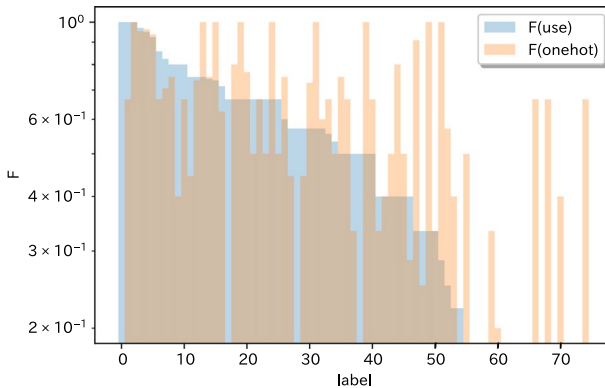
3.3 Results

3.3.1 Quantitative comparison

Table 3 shows the accuracy of each model for the four cases: civil codes, criminal codes, civil court precedents, and criminal court precedents. Scores highlighted in bold font show the best-performing model for the four cases. The hyper-parameter m , which defines the number of additional predictions besides the number of masks, was varied between 0 and 4. Overall, in terms of accuracy, embedding based on

Table 4 F-measure scores

Model	Category	F_{μ}	P_{μ}	R_{μ}	F_M
onehot+XGB	Civil code	7.57	7.18	8.54	3.9
onehot+XGB	Criminal code	46.89	54.52	45.76	24.89
onehot+XGB	Civil case	41.42	47.08	41.75	22.13
onehot+XGB	Criminal case	49.04	48.44	56.41	26.06
USE+XGB	Civil code	13.94	17.55	13.65	7.68
USE+XGB	Criminal code	50.33	58.8	48.31	26.52
USE+XGB	Civil case	41.73	50.84	39.94	22.37
USE+XGB	Criminal case	44.27	43.94	52.38	23.89
T5	Civil code	22.74	25.15	23.68	12.2
T5	Criminal code	21.36	25.57	21.13	11.57
T5	Civil case	7.41	10.11	7.00	4.14
T5	Criminal case	1.38	1.8	1.55	0.83
doc2vec+XGB	Civil code	0.00	0.00	0.03	0.00
doc2vec+XGB	Criminal code	0.33	0.82	0.47	0.30
doc2vec+XGB	Civil case	0.04	0.07	0.09	0.04
doc2vec+XGB	Criminal case	3.23	2.73	4.17	1.65
LDA+XGB	Civil code	4.87	4.43	5.69	2.49
LDA+XGB	Criminal code	28.20	35.22	27.09	15.31
LDA+XGB	Civil case	26.78	29.68	28.72	14.60
LDA+XGB	Criminal case	5.01	4.43	6.33	2.61

**Fig. 7** F-measure scores for criminal cases

word frequency tended to have the highest score, followed by USE. Performance for the civil article is an exception to this pattern, where the T5 shows the best performance.

As in the accuracy case, the T5 model attained the highest prediction score for civil codes for the F-measure. However, in contrast to the accuracy measure, USE tended to have the highest score, followed by word frequency-based embedding

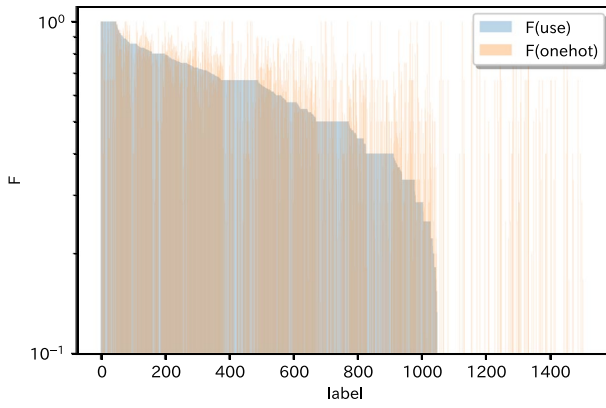


Fig. 8 F-measure scores for civil cases

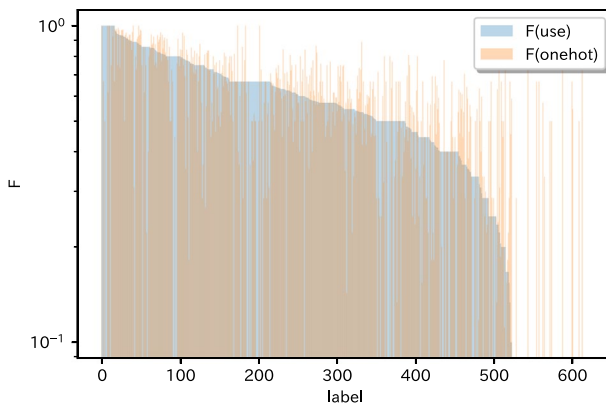


Fig. 9 F-measure scores for criminal codes

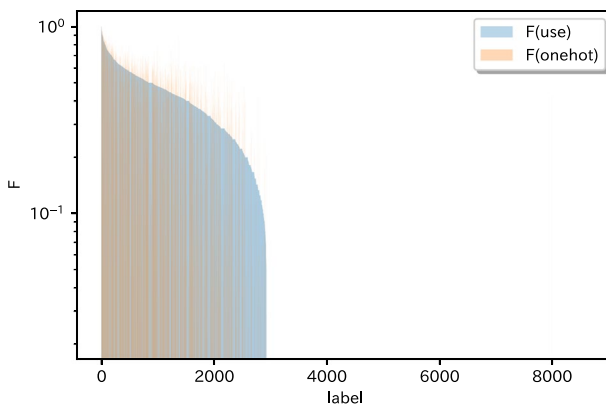


Fig. 10 F-measure scores for civil codes

in the F-measure (Table 4, where scores highlighted in bold font show the best-performing model for the four cases.). The reverse pattern of prediction accuracy between word frequency and USE could be explained by the difference in achieving high prediction among various labels. Figs. 7, 8, 9, 10 shows the F-measure for each label. The vertical axis corresponds to the F-measure for each model (blue shading corresponds to USE, and orange lines to word frequency). The index of the labels was sorted using the number of appearances in the data set from left to right. Hence, labels located in the left part appeared most frequently. As seen in the right part of the figure, neither USE nor word frequency could handle labels with minor appearances. However, USE achieved higher predictive accuracy for more labels compared with the word frequency counterpart. One interpretation of these results is that USE considers the modification relations of words in the order of sentences and is therefore able to grasp the semantics of a sentence better than the word frequency counterpart. Instead, word frequency highlights some keywords, but fails to capture the meaning when the context is more important. We explore this hypothesis in the next qualitative comparison section.

3.4 Qualitative assessment

As already mentioned, since USE is trained on general documents, it assigns a unique token, called Unknown token, to technical legal terms that it has never seen before. This discrepancy makes USE fail to predict cases when the model could quickly pinpoint the law from a keyword. For instance, in the following example: “したがって、被告人につき刑法 197 条 1 項前段の収賄罪の成立を認めた原判断は、正当である。(Therefore, the original judgment that the accused was guilty of bribery under Article 197(1), the first sentence of the Penal Code, is justified.),” the character “賄(wai)” in the word “収賄(shu-wai)” is an unknown token for USE. The term “収賄(shu-wai)” means “receiving a bribe” in Japanese. Article 197 of the Penal Code bans public officials from receiving bribes. Without recognizing “賄(wai)” (bribe), USE is left with only “収(shu)” (receive), making it difficult to pinpoint the statutory law.

In another example: “しかしながら、刑法 62 条 1 項の幫助犯に関する規定は、刑法以外の法令の罪についても、その法令に特別の規定がある場合を除いて適用されるのであり....(However, the provision on aiding and abetting crimes in Article 62(1) of the Penal Code applies to crimes under statutory laws other than the Penal Code unless the rules have special provisions...),” the character “幫(hou)” in the word “幫助(houjyo)” is an unknown token for USE. The term “幫助” means “aiding and abetting.” As in the previous example, since Article 62 of the Penal Code is the law that defines aiding and abetting, and “幫” is the keyword that predicts this law article. Without it, USE outputs the wrong answer.

In addition to the above examples, where USE fails to incorporate some essential keywords, there are other examples where USE fails because it overly focuses on the context. For instance, in the following example, “MS 実験に関するメモ, 新炸薬の開発に関するメモ, 黒色火薬の製造に関するメモ, 新型信管の製造に関するメモの記載内容が真実であることを前提に,...被告人が各実験に立

ち会ったり,黒色火薬等を製造するなどして本件両事件に関与した旨認定し,上記各メモを記載内容の真実性を立証するために用いており,刑訴法 320 条(伝聞法則)に反する。(Based on the premise that the memo's contents about the MS experiment, the note on the development of the new explosive, the note on the manufacture of black powder, and the note on the manufacture of the new fuse are true, ... the above note was used to prove the truthfulness of the contents, which is against Article 320 (hearsay law) of the Code of Criminal Procedure.)”, most of the sentence is about explosives, and the last part briefly mentions the hearsay evidence. The correct answer, in this case, is Article 320 of the Code of Criminal Procedure, which defines what constitutes hearsay evidence. The word frequency-based method answered correctly based on the hearsay evidence. On the other hand, USE incorrectly predicted Article 3 of the Explosive Ordinance (a law that punishes the manufacture of explosives) after being misled by the long sentence.

Contrary to the above examples, where USE performs inferiorly compared with the word frequency counterpart, we next illustrate when USE outperforms the word frequency counterpart. In the following example, the prosecutors wanted to invite a witness who was an accomplice in a related criminal case. This accomplice had already served his sentence, and he tried to avoid an open hearing on a trial date because that would have revealed to the press that he was involved in a major crime. However, the defendants wanted an open hearing because that would have made it easier for the defendants to fight back against the witness. The defendants argued that there was no need for a closed hearing on a date other than the trial date because such worries could be prevented by taking shielding measures to hide the witness from the audience. The exact sentence is as follows: “弁護人は,... 公判回避のために期日外尋問をするのは認められないというのが通説であり,本件で期日外尋問をするかについては慎重に判断すべきである,既に尋問を実施した別の事件では,遮へい措置があれば供述できることが明らかであるから,刑訴法 281 条には該当しない... と意見を述べた。(The defense counsel expressed his opinion that ... it is a standard theory that a closed hearing on a date other than the trial date is not allowed to avoid an ordinary trial, that the court should carefully judge whether to conduct a closed hearing interrogation in this case and that it does not fall under Article 281 of the Code of Criminal Procedure because it is clear that in another case where the interrogation was already conducted, it was possible to make a statement regarding whether there were shielding measures.)”. The word frequency model predicted Article 157-3 of the Code of Criminal Procedure, which defines a provision on shielding measures, possibly dragged by the keyword “shielding measures.” The correct answer, in this case, is Article 281 of the Code of Criminal Procedure, which establishes the provision of an out-of-date examination (i.e., “the court may examine the witness on a day other than the trial date (<http://www.japaneselawtranslation.go.jp/>)”). USE correctly predicted this legal code by reflecting the context of the entire sentence and whether an out-of-date interrogation is inappropriate.

This tendency of the word frequency method to focus only on keywords can be seen in another example concerning self-surrender: “加えて,銃砲刀剣類所持等取締法31条の 5 及び 10 が規定する自首減輕は,けん銃や実包の提出を促してその早期回収を図り,当該けん銃等の使用による危険の発生を極力防止しよ

うという政策的な考慮に基づくものであるところ、Aという経過をたどったことにも照らすと、被告人がけん銃及び実包を提出して自首したということもできないというべきである。(In addition, the reduction of self-surrender provided for in Articles 31-5 and 31-10 of the Act for Controlling the Possession of Firearms or Swords and Other Such Weapons is based on the policy consideration of preventing the occurrence of danger by using such guns by encouraging the submission of firearms and cartridges for early recovery. In light of the fact that the defendant followed the process described in Section A, it should not be possible to say that the defendant surrendered himself by submitting the gun and the actual package.)". The word frequency model predicted Article 42 of the Penal Code, defining general self-surrender. However, the above sentence refers to when the defendant who self-surrendered also submitted a gun to the authorities (and whether his sentencing should be reduced on this basis). This reduction of sentencing with the submission of a firearm is defined in Article 31-5 of the Act for Controlling the Possession of Firearms or Swords and Other Such Weapons, and USE correctly predicted this article by taking this context information into account⁸.

The observation that the word frequency model tends to extract keywords directly connected to a specific legal code or court precedent could be further confirmed by analyzing the feature importance. As the feature importance, we used an average gain in tree split using the features in gradient boosting. For instance, the following keywords were selected as the essential features for the criminal code: “算入 (Counting), 費用 (Cost), 言渡し (Sentence), 重い (Heavy), 併合 (Combined), 言い渡し (Sentence), 加重 (Aggravated), 処断 (Punishment), 上告 (Appeal), and 控訴 (Appeals).” The top important feature, “counting,” is often used to account for days of detention until a court judgment is given. The exact procedure of handling these days of detention is defined in Article 21 of the Penal Code. The word “sentence” is typical in acquittals, citing Article 336 of the Code of Criminal Procedure. The next most significant term, “expenses,” refers to litigation expenses, such as travel expenses for witnesses; whether the accused should owe these expenses is defined in Article 181 of the Code of Criminal Procedure.

Similar observations could be performed for the civil codes where the following words were identified as important features: “損害 (damage), 事故 (accident), 遅延 (delay), とおり (as), 賠償 (compensation), 民事 (civil), 適用 (application), 悪意 (malice), 申出 (offer), and 利息 (interest).” The most frequent civil codes in the court judgments are Article 709 of the Civil Code (“compensation” for “damages” caused by tortious acts), Article 1 of the State Redress Act (tortious acts of public officials and their liability for “compensation,” right of recourse), Article 704 of the Civil Code (obligation of “malicious” beneficiaries to make restitution, etc.), Article 61 of the Code of Civil Procedure (principle of bearing litigation costs), and the

⁸ This law might sound unfamiliar for audiences outside of Japan, so we elaborate on it here. First, gun possession is illegal in Japan, as defined in the firearms law. However, the number of gun-related crimes, such as murders committed by non-gangsters, increased in the early 1990s, causing the National Police Agency to enact a policy to confiscate more guns from the general public. To boost this policy, Article 31-5 of the Act for Controlling the Possession of Firearms or Swords and Other Such Weapons was set, which gave special treatment to self-surrendering with the submission of firearms.

former Article 43 of the Money Lending Business Act (when companies could use the higher “interest” rate). Comparing the top features and the keywords in quotation marks in the previous sentence, we can see why these features were judged as important.

A similar analysis could be performed for court precedents. For criminal court precedents, the following words were judged as important: “構成 (constituent), 予防 (prevention), 選択 (choice), 作成 (create), 意見 (opinion), 対象 (subject), 制度 (system), 提起 (raise), 明確 (clear), and 通常 (normal).” The keywords with the highest contribution included those related to the death penalty, of which there were several court precedents in Japan that justify its constitutionality. The word “constituent” is mainly used in the constituent elements of a crime. “prevention” is primarily used in the idioms “general prevention effect” or “special prevention effect.” While the general preventive effect prevents crimes committed by ordinary citizens, the special preventive effect precludes the defendant from committing crimes again. The sentences that include these phrases cite court precedents that discuss these effects, thus making these keywords predict the correct court precedents. For civil court precedents, “民事判決 (civil judgment), やむを得ない (unavoidable), 酷 (harsh), 入国 (entry), 在留 (stay), 長さ (length), 適合性 (suitability), 所要 (need), 次いで (next), and 続け (continue)” are judged as important. The words “unavoidable” and “harsh” are often used in sentences where there is almost no dispute about the facts, but the law’s applicability is in question.

3.5 Discussion on T5

Contrary to the pattern that gradient boosting methods using word frequency and USE show superior performance for three out of the four data sets, the T5 shows the highest predictive accuracy for the civil code prediction in terms of accuracy and the F-measure. One possible reason for this superior performance is the sheer size of the data set. Since the T5 is a large model, it might require more data sets to train the model correctly. Another possible reason is the tokenizer. Compared with USE, the T5’s tokenizer is extensive, with no unknown token words in the data set. The ability to recognize technical terms appearing in Japanese legal documents correctly might have positively affected the accuracy. Finally, as we noted before, the difference in the loss function of the decoder might have had a positive impact on the predictive performance.

The T5 model is a generative model that sometimes outputs statutory law names that do not exist. For instance, “呼ばわりされた行為の処罰に関する法律5条 (Article 5 of the Act on Punishment of Denouncing)”⁹ and “焼損事件に係る補償に関する法律2条 (Article 2 of the Act on Compensation for Burnout Incident)”¹⁰ are two interesting law names that the T5 model generated but do not exist. The

⁹ This probably came from Article 231 of the Penal Code, which defines insults as a crime.

¹⁰ This resembles the “Act on Compensation for Pollution-related Health Damage,” which is a law that defines compensation from the government for pollution-related health damage.

fact that the off-the-shelf T5 model could output such a realistic law name is quite surprising.

To sum up, USE beat the word frequency counterpart when the statutory laws and court precedents could not be identified based solely on keywords. Moreover, there was some discrepancy in the USE tokenizer, making USE ignore essential keywords related to the law. The T5 model does not have this tokenizer problem, but requires more data to train the decoder than does the gradient boosting method because of its large model size. However, when we have a large data set (as in the case of civil code prediction), it outperforms the other models owing to the different loss functions employed in the decoder. These observations suggest the need for building deep learning models specifically adapted to the legal domain, as in the former research of Chalkidis et al. (2020) and Tagarelli and Simeri (2021).

4 Legal link prediction

When a judge concludes a case, their judgment is not only driven by one legal code or court precedent. Usually, there are several key issues that a judge has to decide, each of which requires a reexamination of the interpretation of the statutory laws and court precedents. Moreover, when there are no precedents for the case in question, judges need to develop new arguments that might, in turn, be cited by future court judgments.

To give a concrete example of how multiple legal codes and court precedents interact and evolve, during the 2010s, the legality of GPS investigations without a warrant became a controversial issue in Japan. The central controversy was whether (i) GPS investigations without a warrant are, in fact, legal, and (ii) if they are illegal, whether evidence obtained through GPS investigations can be used as evidence in a criminal trial. Numerous codes (e.g., Article 197(1) of the Code of Criminal Procedures, Article 35(1) of the Constitution) and court precedents (e.g., Supreme Court, September 7, 1978, Supreme Court, February 14, 2003) have established the exclusion principle of legal investigation and illegally collected evidence. Still, it was not enough to decide fully how to deal with this new investigation technology (i.e., GPS devices).

Even though the high court stated that evidence collected through GPS investigations without a warrant could be included in a criminal trial, the supreme court overturned the decision, ruling that the GPS evidence in the current trial should be dismissed. Moreover, the supreme court suggested that a new statutory law that defines the condition and procedure of such GPS searches should be legislated (Judgment of the grand bench of 2016 2017). There are several things worth noting from this example. First, this new court precedent connects many legal codes and court precedents in criminal procedure, also filling in the gap that the new technology created in the current legal system. Moreover, until the law defining the condition and procedure of GPS investigations is legislated in the future, this court precedent would be

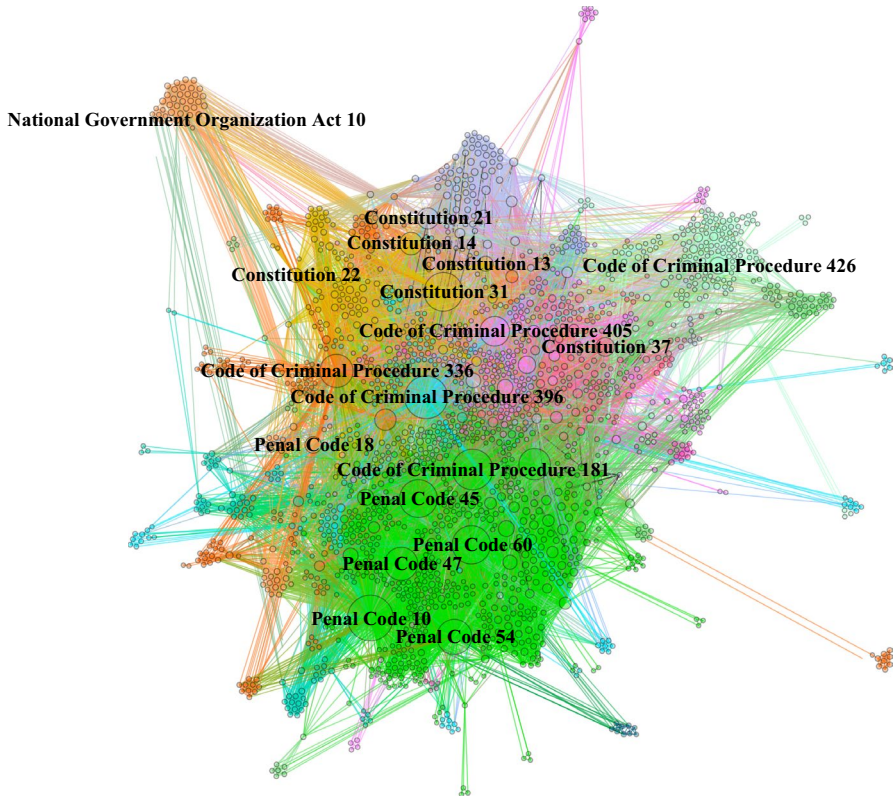


Fig. 11 Network visualization of a criminal case

the defining precedent that would make it nearly impossible to conduct GPS investigations in Japan¹¹. As can be understood from this example, statutory laws and court precedents interact with each other and sometimes evolve with new cases brought into the courtroom¹².

Another important point about court judgments is that one court judgment might not necessarily provide the entire logic behind a decision. For instance, the court precedent of the Supreme Court from June 13, 2006 (Judgment of the second petty bench of 2004 2006) was a decisive judgment that gave the interpretation of the former Article 43 of the Money Lending Business Act. As described above, this was a massive game-changer in the Japanese legal profession and consumer finance companies. However, many following court judgments took these precedents for granted and sometimes even omitted to mention them. The most cited case related to excessive loan payments was the court precedent of the Supreme Court from July

¹¹ At the time of writing, this law has not been legislated, making GPS investigations nearly impossible.

¹² The evolutionary view of the legal system through the juridical process is seen in many great writings of the past, keeping a common law system in mind, as well. See (Tamanaha 2017; Tamanaha and Z 2004; Hayek 1973; Smith et al. 1982) for details.

Table 5 Page rank of criminal cases

Rank	Law name	Page rank
1	Penal Code 10	0.015
2	Code of Criminal Procedure 396	0.014
3	Code of Criminal Procedure 181	0.013
4	Constitution 31	0.013
5	Penal Code 60	0.013
6	Penal Code 45	0.012
7	Penal Code 54	0.011
8	Code of Criminal Procedure 336	0.011
9	Penal Code 47	0.011
10	Penal Code 21	0.01

13, 2007 (Judgment of the second petty bench of 2005 2007), which decided that excessive interest should be refinanced to customers with a statutory interest rate (which was 5 percent a year), as shown in Fig. 2. This example shows that using the co-occurrence pattern of a single court judgment might not describe the entire picture of how the law is used. Building a network by patching co-occurrence patterns enables us to grasp the whole picture.

Fig.11 shows the aggregated network of co-occurrence patterns in the entire data set for the criminal cases. We ignored all the edge weights and determined the position of each node by using Force Atlas 2 (Jacomy et al. 2014). The colors show the communities identified by the standard modularity maximizing algorithm. We could see that the Code of Criminal Procedures clustered at the lower right along with important past court precedents. Table 5 also shows the top-ranking nodes in a simple centrality measure (i.e., Page rank Page et al. 1998). We could see that Article 31 of the Constitution, which describes the no punishment without law principle, is listed at the top of the Page Rank score. This is in-line with intuition because no punishment without the law (“*zaikai-hotei-shugi*” in Japanese) is the defining principle that enables the state to penalize individuals. These are simple observations, but prove that creating a network in this fashion provides meaningful insights into the inner working mechanisms of the law’s interdependence structure in the Japanese legal system.

There are several ways to define a link prediction problem using this network. One is a static network link prediction problem where we randomly omit a certain number of edges from the network and predict link probability. Although this is also an exciting topic, we argue that defining a link prediction problem in a dynamic setting where we split the network by the published date of the underlying court judgments and predict the future combinations that did not appear in the past is more interesting. This legal link prediction task corresponds to the case where we try to predict the interaction of law that is plausible from a network perspective, but has not yet occurred because no claims were brought into court that could give rise to such interactions. Thus, we perform link prediction in the dynamic setting.

We use the co-occurrence networks before January 1, 2010, as our training data set and the rest as our test data set. This setting roughly eliminates 10 percent of the positive edges from the data set. Although performing a link prediction problem in this setting ignores all the nodes (statutory laws and court precedents) that never appeared before, it still captures the above motivation. The way to incorporate these unseen nodes can be divided into two categories. One uses all the textual information of all the nodes (statutory laws and court precedents) and builds a model that can perform a prediction without seeing the link patterns of some nodes in the training set. The other involves creating a different legal network using the citation information written in the legal codes and building a multiplex network. Although these are interesting topics, the amount of work needed to construct such a data set is massive, so we left this for future work.

4.1 Models and evaluation metrics

We compared the following models, from simple network models from complex network literature to the more advanced deep learning methods found in the machine learning literature.

- Adamic-Adar is an index proposed in Adamic and Adar (2003). It evaluates the likelihood of a link based on common neighbors shared between nodes. Specifically, it is defined as $A(x, y) = \sum_{u \in N(x) \cap N(y)} \frac{1}{\log |N(u)|}$, where $N(u)$ is the set of adjacent nodes of u .
- The Jaccard coefficient is a coefficient similar to Adamic-Adar. The difference is in the normalization step, as specified in the following equation:
- Preferential attachment is yet another basic link prediction score. It is defined by the following equation: $|N(u)||N(v)|$.
- The Stochastic Block Model (SBM) (Holland et al. 1983) is a canonical latent block model that assumes that nodes are assigned to a block, and the interaction probabilities among the blocks fully determine the likelihood of a link. The limitation of the SBM is that theoretically, it is a model suited for no heavy-tailed distributions.
- The degree-corrected stochastic block model (DCSBM) (Karrer and Newman 2011) is a variant of the SBM that incorporates degree heterogeneity. It is well known that from a model selection perspective, the DSCBM is often preferred over the SBM when the node degree distribution is heavy-tailed.
- Node2vec is a method that embeds nodes to vectors using context information defined by random walks (Grover and Leskovec 2016).
- Attri2vec is a method that adjusts for node attributes using node2vec to ensure structural similarity (Zhang 2019).
- The graph convolutional network (GCN) is the basic graph neural network model where node embeddings are calculated via graph convolution (Kipf and Welling 2017).

Table 6 Link prediction results. Text feature indicates whether textual information was used in the model. ROC stands for the AUC-ROC score and PR for the AUC-PR score. Bold fonts represent top scores

Model	Text feature	Criminal ROC	Criminal PR	Civil ROC	Civil PR
Adamic-Adar	No	0.850	0.634	0.849	0.662
Resource allocation	No	0.859	0.668	0.856	0.686
Jaccard coefficient	No	0.743	0.310	0.761	0.391
Preferential attachment	No	0.853	0.596	0.823	0.570
SBM	No	0.880	0.678	0.852	0.645
DCSBM	No	0.910	0.722	0.873	0.679
node2vec	No	0.718	0.346	0.730	0.396
att2vec	Yes	0.640	0.640	0.583	0.231
GCN	Yes	0.910	0.728	0.901	0.716

We used the area under the receiver operating characteristic curve (AUC-ROC) and area under the precision-recall curve (AUC-PR) evaluation metrics. These are standard performance metrics used in the link prediction problem that evaluate link prediction performance over the average of all thresholds.

4.2 Results

4.2.1 Quantitative comparison

Table 6 summarizes the results. We could see that Adamic-Adar, Jaccard coefficient, and preferential attachment, which are quite simple measures, already perform well in criminal and civil cases. However, the SBM and DCSBM, which consider the mesoscale block structure, perform even better. The DCSBM performs better than the SBM counterpart, which may be attributed to the heavy-tailed degree distribution. For the deep learning model, the simple node2vec underperforms even the simple prediction of Adamic-Adar. Although attr2vec, which simplistically utilizes text information, showed the worst performance, the GCN, which also considers textual information, scored the best in terms of criminal and civil cases. This quantitative analysis shows that even a simple model achieves fairly good performance in dynamic link prediction. However, a sophisticated model that incorporates both network and textual information performs best, which highlights the need to create a model that can learn from the two data sources.

4.2.2 Qualitative comparison

We give further insights into the learned model for the GCN, which showed the best performance for the models compared here. We took the learned embeddings for each law and court precedent and reduced its dimension to two using t-SNE (van der Maaten and Hinton 2008). Fig. 12 shows the results for the civil case, and Fig. 13 for the criminal case. We see that the law and court precedents tended to be separated

Fig. 12 t-SNE visualization of node embedding (civil)

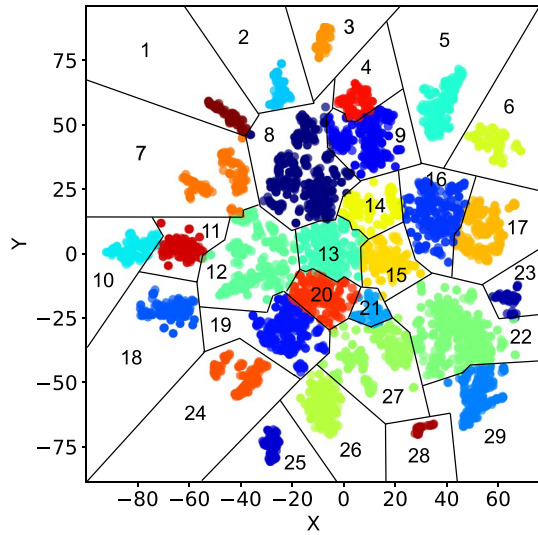
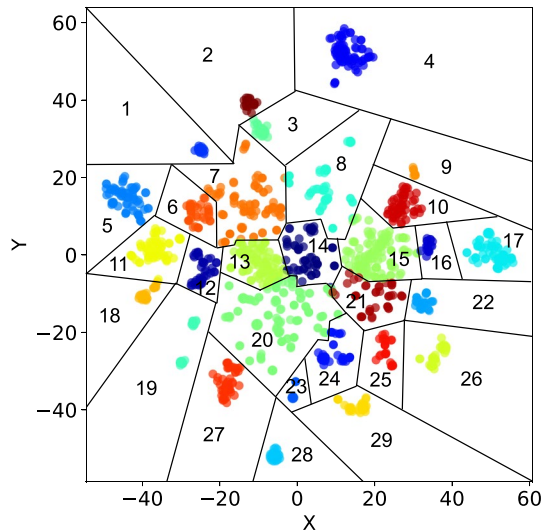


Fig. 13 t-SNE visualization of node embedding (criminal)



into distinct clusters. Using this insight, we performed clustering using the Gaussian mixture model, of which the result is illustrated by colors (or numbers with unique regions). We see 29 distinct groups for the civil cases and 28 for the criminal cases.

Civil cases cover a wide range of the law because disputes among individuals or institutions can take various forms. For example, we can see that cluster 18 in Table 7 covers issues concerning trademarks and unfair competition. Both the Trademark Act and Unfair Competition Prevention Act are necessary to stop the use of companies' logos or brands and to claim damages upon illegal usage. Therefore, they are often used together and can be broadly lumped under "intellectual property

Table 7 Clustering results for the civil cases

Cluster	Statutory laws and court precedents
1	Public Offices Election Act; Local Autonomy Act; Constitution 42,43,44,45,46,47,54,59; House of Representatives Election District Election Council Establishment Act; Court precedents
2	Immigration Control and Refugee Recognition Act; Civil Code 752,809,877; Nationality Act 11; Court precedents
3	Constitution 9,17,40,41,51,81,98; Civil Code 11,16,158,174,184,185,188,194,373,393,494, 723; State Redress Act 6; Court precedents
4	Local Civil Service Act; Local Autonomy Act; Local Government Finance Act; Basic Act on Education 10; Constitution 23; Court precedents
5	Labor Standards Act; Labor Contract Act; Labor Union Act; Employment Insurance Act; Court precedents
6	Pneumoconiosis Act; Labor Standards Act; Industrial Safety and Health Act; Industrial Accident Compensation Insurance Act; Ordinance on Prevention of Hazards Due to Specified Chemical Substances; Court precedents
7	Basic Resident Registration Act; Act on the Protection of Personal Information; Act on Access to Information Held by Administrative Organs; Court precedents
8	Constitution; Child Welfare Act; National Pension Act; Employees' Pension Insurance Act; Nationality Act; School Education Act; Court of Civil Procedure; Civil Code; Court precedents;
9	National Public Service Act; Local Public Service Act; Local Public Enterprise Act; Act on the Organization and Operation of Local Educational Administration; Local Autonomy Act; Constitution 39,94; Court precedents
10	Patent Act; Utility Model Act; Design Act; Court precedents
11	Patent Act; Utility Model Act; Design Act; Order for Enforcement of the Act on Securing Quality, Efficacy and Safety of Products Including Pharmaceuticals and Medical Devices; Court precedents
12	Copyright Act; Act on Prohibition of Private Monopolization and Maintenance of Fair Trade; Code of Civil Procedure; Civil Code; Court precedents;
13	Commercial Code; Civil Code; Court of Civil Procedure; Industrial Accident Compensation Insurance Act; Court precedents
14	Medical Practitioners Act; Atomic Bomb Survivors' Assistance Act; Civil Code; Court precedents
15	Road Traffic Act; Act on Securing Compensation for Automobile Accidents; Civil Code 153,404,405,419,493,712,713,714,718,884; Code of Civil Procedure 117, 260; Civil Rehabilitation Act 87; Court precedents
16	Waste Management and Public Cleansing Act; Administrative Case Litigation Act; Administrative Complaint Review Act; Health Insurance Act; Land Readjustment Act; Court precedents
17	Building Standards Act; Expropriation of Land Act; River Act; Basic Environment Act; City Planning Act; Court precedents
18	Trademark Act; Unfair Competition Prevention Act; Court precedents
19	Financial Instruments and Exchange Act; Trust Act; Commercial Code; Civil Code; Companies Act; Banking Act; Court precedents

Table 7 (continued)

Cluster	Statutory laws and court precedents
20	Civil Code; Commercial Code; Financial Instruments and Exchange Act; Court precedents
21	Real Property Registration Order; Civil Code; Court precedents
22	Act on General Rules for National Taxes; Local Tax Act; Income Tax Act; Consumption Tax Act; Inheritance Tax Act; Civil Code; Court precedents
23	Local Tax Act; Cropland Act 4,5; Court precedents
24	Money Lending Business Act; Interest Rate Restriction Act; Civil Code; Code of Civil Procedure; Court precedents
25	Companies Act; Civil Provisional Remedies Act 7,13,23; Court of Civil Procedure 188,331; Court precedents
26	Bankruptcy Act; Civil Rehabilitation Act; Act on General Rules for National Taxes; Commercial Code; Court precedents
27	Act on Land and Building Leases; Act on Building Unit Ownership; Civil Code; Inheritance Tax Act; Civil Execution Act; Court precedents
28	Income Tax Act 9,30,34; Court precedents
29	Act on General Rules for National Taxes; Income Tax Law; Corporation Tax Act; Consumption Tax Act; Act on Special Measures Concerning Taxation; Court precedents

law,” making it natural for them to be categorized in the same cluster. Specifically, Article 4 of the Trademark Act, included in this cluster, defines trademarks that cannot be registered under Japanese law. Numerous court precedents have provided specific interpretations of this article and its application to actual cases. To give a particular example, the Supreme Court precedent of September 8, 2008 (also included in cluster 18), made a critical judgment on a case regarding whether a trademark is “similar” under Article 4 of the Trademark Act (Article 4(1)(11) to be precise). In establishing the criteria for what “similarity” means, the court precedent cited the precedents of the Supreme Court of February 27, 1968 and December 5, 1963, which gave significant decisions (i.e., or “*kihan-teiritsu*” in Japanese) and also happened to belong to the same cluster. This clustering makes visible how statutory laws and court precedents are used in conjunction in real-world trademark cases.

Clusters 10 and 11, adjacent to cluster 18, contain many articles from the Patent Act, Utility Model Act, and Design Act, and cluster 12 contains many articles from the Copyright Act. The aforementioned “intellectual property laws” are often classified as mainly including the Patent Act, Utility Model Act, Design Act, Trademark Act, Copyright Act, and Unfair Competition Prevention Act, and it can be said that the relationships not only within each cluster but also between clusters are well extracted from the GCN result.

Other such examples can be found. For example, cluster 29 relates to tax cases. In particular, it contains many articles of the National Tax Act, Income Tax Act, Corporation Tax Act, and Consumption Tax Act. In addition, many court precedents related to tax cases are also included in cluster 29, Article 30 of the Constitution, which stipulates the tax obligations of citizens, and Article 84 of the Constitution, which sets forth the principle of no taxation without statutory law. It can also be

Table 8 Clustering results for the criminal cases

Cluster	Statutory laws and court precedents
1	Constitution 1,3; Penal Code 1; Code of Criminal Procedure 256; Court precedents
2	Public Offices Election Act; Constitution 12; Police Act 2,26; Court precedents;
3	Constitution 15,16,19,41,73,98; National Public Service Act 19,102,108,110; Code of Criminal Procedure 39; Court precedents
4	Firearms and Swords Control Act 31,32; Anti-Prostitution Act; Child Welfare Act; Act on Punishment of Organized Crimes and Control of Proceeds of Crime; Narcotics and Psychotropics Control Act; Act on Punishment of Physical Violence; Court precedents
5	Juvenile Act; Offenders Rehabilitation Act 50,68,72; Child Welfare Act 5,25,27; Constitution 40; Criminal Compensation Act; Code of Criminal Procedures; Penal Code 134,225; Court precedents
6	Code of Criminal Procedure; Constitution 33; Court precedents
7	Code of Criminal Procedure; Constitution 7,9,11,13,14,18,20,21,31,32,35,36,37,38, 39,65,76,82,97; Court precedents
8	Labor Standards Act; Health Insurance Act 13; Court precedents
9	Penal Code 10,21,45,47,54,60; Code of Criminal Procedure 181,396
10	Penal Code; Act on Punishment of Organized Crimes and Control 3,13; Firearms and Swords Control Act 3
11	Code of Criminal Procedures; Constitution 34,81; Court precedents
12	Code of Criminal Procedures; Measurement Act 17; Court precedents;
13	Code of Criminal Procedures; Child Welfare Act 7; Constitution 79; Road Traffic Act 52; Court precedents
14	Dental Practitioners Act 28,31; Rabies Prevention Act 5,6; Medical Practitioners Act 17,31; Industrial Safety and Health Act 30; Penal Code; Court precedents
15	Penal Codes; Act on Prohibition of Possession of Special Picking Tools, and Other Related Matters 4; Act on Punishment of Organized Crimes and Control of Proceeds of Crime 2,4,13,16; Organ Transplant Act 2; Narcotics and Psychotropics Control Act 1,65; Court precedents
16	Act Concerning Special Provisions for the Narcotics; Court precedents
17	Road Traffic Act; Road Transport Vehicle Act; Court precedents
18	Code of Criminal Procedures 188,281,299,316; Attorney Act 23; Court precedents
19	Code of Criminal Procedures 191,195; Penal Code 193; Public Prosecutor's Office Act 4,5; Act on Preventing Collisions at Sea 5; Court precedents
20	Labor Standards Act 9,20,21,22; Industrial Safety and Health Act; Small and Medium Enterprises Basic Act 3; Medical Practitioners Act; Constitution 17; Court precedents
21	Act on Prohibition of Unauthorized Computer Access 3,8; Unfair Competition Prevention Act; National Public Service Act 100,109; Political Funds Control Act 12,25; Court precedents
22	Income Tax Act; Corporate Tax Act; Court precedents
23	Smaller Enterprise Retirement Allowance Mutual Aid Act; Public Accounting Act;

Table 8 (continued)

Cluster	Statutory laws and court precedents
24	Public Accounting Act; Cabinet Order on Budgets, the Settlement of Accounts and Accounting; Local Autonomy Act; Court precedents
25	Customs Act; Consumption Tax Act 64; Penal Code 18; Court precedents
26	Banking Act; Commercial Code; Court precedents
27	Waste Management and Public Cleansing Act; Tobacco Business Act 1,2; Basic Act for Establishing a Recycling-based Society 3,10; Constitution 22,25,94; Court precedents
28	Diet Act; National Government Organization Act; Act on the Organization and Operation of Local Educational Administration
29	Foreign Exchange and Foreign Trade Act; Waste Management and Public Cleansing Act; Customs Act 67,111,113,117; Court precedents

seen that clusters 22 and 28, which surround cluster 29, contain many articles and rulings about tax laws.

The number of articles and judgments in criminal cases is far smaller than that in civil cases, while the diversity of claims brought into court is less than that in civil cases. This makes it difficult to state clearly the characteristics of each cluster compared with the civil case counterpart. However, cluster 17 in Table 8, for example, includes many articles from the Road Traffic Act and Road Transport Vehicle Act, indicating that criminal laws and court precedents related to public transport safety form a cluster slightly different from the other clusters.

In addition, laws related to the protection of juveniles and children, such as the Juvenile Act and Child Welfare Act, are grouped in cluster 5. Furthermore, Articles 10, 45, 54, and 60 of the Criminal Code, as well as Articles 396 and 181 of the Code of Criminal Procedure, which appeared at the top in Table 5, all belong to cluster 9. These are nodes with a reasonably high degree in the network since they often appear simultaneously and are cited a large number of times. It is noteworthy that these frequently appearing nodes are grouped into a single cluster, even though the frequency of occurrence and the weight of the edges are not taken into account in the current analysis.

5 Conclusion

In the present paper, to data-mine how statutory laws and past court precedents are used and how their usage has evolved in the courtroom, we proposed two tasks to capture such a structure and compared the strengths and weaknesses of major machine learning models. One is a prediction task based on masked language modeling where the goal is to predict the statutory laws and court precedents' names from the sentence masking them. The other is a dynamic link prediction task where the goal is to identify novel interactions among the statutory laws and court precedents

that have not occurred in the past. We then performed a quantitative comparison among the major machine learning models and provided a detailed qualitative comparison of the learned structure. The insights from the current paper motivate further developments in legal data science.

There are many avenues for future works. For the language model, our setting in the current paper is an extreme class classification problem where there are many labels to be predicted from the text information alone. Our model (especially the USE model) successfully predicted a law name that appears less frequently than the word frequency counterpart. However, there is still room for improvement because, for some cases (civil articles), we were only able to predict 34 percent of the law names. Using additional information, such as the textual content of statutory laws and court precedents, or using citation network information inside the legal code, might fill this gap and should be explored in future work. Although the gradient boosting model using the world frequency feature was able to make predictions reasonably well by focusing on keywords, the model using USE seemed to be able to grasp the context information from the reason for citation. Future work should incorporate both of these aspects. Building deep learning models tailored for the legal domain based on the findings in this paper is also exciting work left for future research.

The dynamic network that we built in Section 5 is a hypergraph. There are existing models that can embed hypergraphs similar to the ones we performed in this paper. Developing a hypergraph model that suits the legal dynamic link prediction setting is also an exciting task left for future work. Moreover, the clustering result in the link prediction section is limited because we performed hard clustering without considering that one law could be used in different settings (i.e., have multiple meanings). Capturing this “polysemy” of the statutory laws and court precedents is also exciting and left for future work.

Acknowledgements We wish to thank Kazuyuki Yamasawa, Fumihiro Yamawaki, Yutaka Ishikawa, and Katsutoshi Shibagaki of TKC Corporation for the fruitful discussion concerning this work. We would also like to thank Tsutomu Watanabe, Hiroshi Iyetomi, Takaaki Ohnishi, Takayuki Mizuno, Shohei Doi, Tomohiro Watanabe, Tatsuya Nagasawa, Shu Takahashi and the members of the Cannon Institute for Global Studies for helpful discussions concerning this work. This work was supported by JST FOREST Program (Grant Number JPMJFR216Q, Japan) and the UTEC-UTokyo FSI Research Grant Program. The Forte Science Communications polished up the English language.

Author contributions All authors designed the research, performed the research, analyzed the data, and wrote the paper.

Funding This work was supported by JST FOREST Program (Grant Number JPMJFR216Q, Japan) and the UTEC-UTokyo FSI Research Grant Program.

Declarations

Conflict of interest The authors have no competing interests to declare that are relevant to the content of this article.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative

Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Adamic E, Adar LA (2003) Friends and neighbors on the web (3):211–230
- Alex L, Haven L, Michel S, Slater D (2020) Policy disruption early warning system. Google Patents. US Patent 10,776,891
- Badawi AB, Dari-Mattiacci G (2019) Reference networks and civil codes. See Livermore Rockmore 2019:339–365
- Blei DM, Ng AY, Jordan MI (2003) Latent dirichlet allocation. *J Mach Learn Res* 3:993–1022
- Boulet R, Mazzega P, Bourcier D (2011) A network approach to the french system of legal codes-part i: analysis of a dense network. *Artif Intell Law* 19(4):333–355
- Boulet R, Mazzega P, Bourcier D (2018) Network approach to the french system of legal codes part ii: the role of the weights in a network. *Artif Intell Law* 26(1):23–47
- Chalkidis I, Kampas D (2019) Deep learning in law: early adaptation and legal word embeddings trained on large corpora. *Artif Intell Law* 27(2):171–198
- Chalkidis I, Fergadiotis M, Malakasiotis P, Aletras N, Androutsopoulos I (2020) Legal-bert: The muppets straight out of law school. arXiv preprint [arXiv:2010.02559](https://arxiv.org/abs/2010.02559)
- Chen T, Guestrin C (2016) Xgboost: A scalable tree boosting system. In: Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining, pp. 785–794
- Coupette C, Beckedorf J, Hartung D, Bommarito M, Katz DM (2021) Measuring law over time: a network analytical framework with an application to statutes and regulations in the united states and germany. *Front Phys* 9:269
- Dadgostari F, Guim M, Beling PA, Livermore MA, Rockmore DN (2020) Modeling law search as prediction. *Artif Intell Law* 29:3–34
- Fowler JH, Jeon S (2008) The authority of supreme court precedent. *Soc Netw* 30(1):16–30
- Grand bench of the supreme court of 1973 (a) no. 910 of september 10, 1975, keishu vol.29, no.8, at 489
- Grover A, Leskovec J (2016) node2vec scalable feature learning for networks. Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining, pp 855–864
- Hayek FA (1973) Law, legislation and liberty, volume 1: Rules and order
- Holland P, Laskey KB, Leinhardt S (1983) Stochastic blockmodels: first steps. *Soc Netw* 5:109–137
- Hustler magazine, inc. v. falwell, 485 u.s. 46 (1988)
- Jacomy M, Venturini T, Heymann S, Bastian M (2014) Forceatlas2, a continuous graph layout algorithm for handy network visualization designed for the gephi software. *PLoS ONE* 9(6):1–12. <https://doi.org/10.1371/journal.pone.0098679>
- Japanese law translation. <http://www.japaneselawtranslation.go.jp/>
- Jones CPA (2020) The Japanese legal system in a nutshell. The Japanese legal system. West Academic, St. Paul, Minn., United States
- Judgment of the second petty bench of 2004 (ju) 1518 of january 13, 2006, minshu vol. 60, no. 1
- Judgment of the second petty bench of 2005 (ju) 1970 of july 13, 2007, minshu vol. 61, no. 5
- Judgment of the grand bench of 2016 (a) 442 of march 15, 2017, keishu vol. 71, no. 3
- Karrer B, Newman MEJ (2011) Stochastic blockmodels and community structure in networks. *Phys Rev E* 83:016107
- Kipf TN, Welling M (2017) Semi-supervised classification with graph convolutional networks. ICLR
- Koniaris M, Anagnostopoulos I, Vassiliou Y (2018) Network analysis in the legal domain: a complex model for European union legal sources. *J Complex Netw* 6(2):243–268
- La Cava L, Simeri A, Tagarelli A (2021) The Italian civil code network analysis
- Le Q, Mikolov T (2014) Distributed representations of sentences and documents. In: International Conference on Machine Learning, pp. 1188–1196. PMLR

- Liu Y-H, Chen Y-L, Ho W-L (2015) Predicting associated statutes for legal problems. *Inf Process Manag* 51(1):194–211
- Lyte A, Slater D, Michel S (2015) Network measures of the united states code. Technical report, MITRE CORP MCLEAN VA
- van der Maaten L, Hinton G (2008) Visualizing data using t-SNE. *J Mach Learn Res* 9:2579–2605
- Mazzega P, Bourcier D, Boulet R (2009) The network of french legal codes. In: *Proceedings of the 12th International Conference on Artificial Intelligence and Law*, pp. 236–237
- Medvedeva M, Vols M, Wieling M (2020) Using machine learning to predict decisions of the european court of human rights. *Artif Intell Law* 28(2):237–266
- Morimoto A, Kubo D, Sato M, Shindo H, Matsumoto Y (2017) Legal question answering system using neural attention. COLIEE@ ICAIL 2017:79–89
- Nanda R, Adebayo KJ, Di Caro L, Boella G, Robaldo L (2017) Legal information retrieval using topic clustering and neural networks. In: COLIEE@ ICAIL, pp. 68–78
- Nguyen T-S, Nguyen L-M, Tojo S, Satoh K, Shimazu A (2018) Recurrent neural network-based models for recognizing requisite and effectuation parts in legal texts. *Artif Intell Law* 26(2):169–199
- Nguyen H-T, Nguyen PM, Vuong T-H-Y, Bui QM, Nguyen CM, Dang BT, Tran V, Nguyen ML, Satoh K (2021) Jnlp team: Deep learning approaches for legal processing tasks in coliee 2021. arXiv preprint [arXiv:2106.13405](https://arxiv.org/abs/2106.13405)
- Page L, Brin S, Motwani R, Winograd T (1998) The pagerank citation ranking: Bringing order to the web. In: *Proceedings of the 7th International World Wide Web Conference*, Brisbane, Australia, pp. 161–172. <http://citeseer.nj.nec.com/page98pagerank.html>
- Raffel C, Shazeer N, Roberts A, Lee K, Narang S, Matena M, Zhou Y, Li W, Liu PJ (2019) Exploring the limits of transfer learning with a unified text-to-text transformer. arXiv preprint [arXiv:1910.10683](https://arxiv.org/abs/1910.10683)
- Sadeghian A, Sundaram L, Wang DZ, Hamilton WF, Branting K, Pfeifer C (2018) Automatic semantic edge labeling over legal citation graphs. *Artif Intell Law* 26(2):127–144
- Sakhaee N, Wilson MC (2021) Information extraction framework to build legislation network. *Artif Intell Law* 29(1):35–58
- Shulayeva O, Siddharthan A, Wyner A (2017) Recognizing cited facts and principles in legal judgements. *Artif Intell Law* 25(1):107–126
- Smith A, Meek RL, Raphael DD, Stein P (1982) *The Glasgow Edition of the Works and Correspondence of Adam Smith: Lectures on Jurisprudence*. The Glasgow Edition of the Works and Correspondence of Adam Smith. Liberty Fund. <https://books.google.co.jp/books?id=YTmgngEACAAJ>
- Sulea O-M, Zampieri M, Malmasi S, Vela M, Dinu LP, Van Genabith J (2017) Exploring the use of text classification in the legal domain. arXiv preprint [arXiv:1710.09306](https://arxiv.org/abs/1710.09306)
- Tagarelli A, Simeri A (2021) Unsupervised law article mining based on deep pre-trained language representation models with application to the italian civil code. *Artif Intell Law*, 1–57
- Tamanaha BZ (2017) *A realistic theory of law*. Cambridge University Press, Cambridge
- Tamanaha BZ, Z TB (2004) *On the rule of law: history, politics, theory*
- Wang P, Yang Z, Niu S, Zhang Y, Zhang L, Niu S (2018) Modeling dynamic pairwise attention for crime classification over legal articles. In: *the 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, pp. 485–494
- Yamakoshi T, Komamizu T, Ogawa Y, Toyama K (2019) Japanese mistakable legal term correction using infrequency-aware bert classifier. In: *2019 IEEE International Conference on Big Data (Big Data)*, pp. 4342–4351. IEEE
- Yang Y, Cer D, Ahmad A, Guo M, Law J, Constant N, Abrego GH, Yuan S, Tar C, Sung Y-H, et al. (2019) Multilingual universal sentence encoder for semantic retrieval. arXiv preprint [arXiv:1907.04307](https://arxiv.org/abs/1907.04307)
- Yoshioka M, Aoki Y, Suzuki Y (2021) Bert-based ensemble methods with data augmentation for legal textual entailment in coliee statute law task. In: *Proceedings of the Eighteenth International Conference on Artificial Intelligence and Law*, pp. 278–284
- Zhang D, Yin J, Zhu X, Zhang C (2019) Attributed network embedding via subspace discovery. *Data Min Knowl Discov*. 33: 1953
- Zhang P, Koppaka L (2007) Semantics-based legal citation network, 123–130