LawyerPAN: A Proficiency Assessment Network for Trial Lawyers

Yanqing An[1], Qi Liu[1,*], Han Wu[1], Kai Zhang[1], Linan Yue[1], Mingyue Cheng[1], Hongke Zhao[2], Enhong Chen[1]
[1]Anhui Province Key Laboratory of Big Data Analysis and Application, School of Data Science & School of Computer Science and Technology, University of Science and Technology of China,
{anyq, wuhanhan, sa517494, lnyue, mycheng}@mail.ustc.edu.cn, {qiliuql, cheneh}@ustc.edu.cn;
[2]College of Management and Economics, Tianjin University, {hongke}@tju.edu.cn

ABSTRACT

Assessing the proficiency of trial lawyers in different legal fields is of significant importance since a qualified lawyer or lawyer team can strive for his clients' best rights while ensuring the fairness of litigations. However, proficiency assessment for lawyers is very challenging due to many technical and domain challenges, such as the lack of unified evaluation standards, and the complex interactions between lawyers and cases in real legal systems. To this end, we propose a novel proficiency assessment network for trial lawyers (LawyerPAN) to quantify lawyer proficiency through online litigation records. Specifically, we first leverage the theories in psychological measurement for mapping the proficiency of lawyers in each field into a unified real number space. Meanwhile, the characteristics of cases (i.e., case difficulty and discrimination) are well modeled to ensure fairness when assessing lawyers in different cases and fields. Then, we model the interactions between lawyers and cases from two perspectives: the anticipatory perspective aims to measure the personal proficiency of anticipated strategy, and the adversarial perspective seeks to depict the gap of lawyers' proficiency between both sides (i.e., plaintiffs and defendants). Finally, we conduct extensive experiments on real-world data, and the results show the effectiveness and interpretability of our approaches on assessing the proficiency of trial lawyers.

CCS CONCEPTS

• **Computing methodologies → Neural networks**; • **Applied computing → Law**.

KEYWORDS

Lawyer Proficiency Assessment; Legal Intelligence; Neural Network

ACM Reference Format:
Yanqing An, Qi Liu, Han Wu, Kai Zhang, Linan Yue, Mingyue Cheng, Hongke Zhao, Enhong Chen. 2021. LawyerPAN: A Proficiency Assessment Network for Trial Lawyers. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '21), August 14–18, 2021, Virtual Event, Singapore.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3447548.3467218

* Corresponding Author.

KDD '21, August 14–18, 2021, Virtual Event, Singapore.

ACM ISBN 978-1-4503-8332-5/21/08…$15.00
https://doi.org/10.1145/3447548.3467218

1 INTRODUCTION

As an emerging trend, more and more legal platforms [1] come to provide online services [33] to meet the increasing legislative requirements, where people can seek legal assistance from lawyers conveniently. However, there are many lawyers who have tremendously different proficiency levels and differentiated areas of expertise, which brings enormous difficulties in choosing appropriate legal teams to the general public who are often unfamiliar with this legal field. Therefore, it is quite necessary to make professional profiles for lawyers by assessing their fine-grained expertise proficiency in various fields [8].

However, the proficiency assessment of trial lawyers is quite a complicated problem. In fact, good lawyers strive for the best rights of their clients, while ensuring the fairness of the proceedings, instead of defining "win" and "loss" the same way most nonprofessional people do [15, 22, 27]. For better understanding, Figure 1 gives a toy example. Tracing back the career of one lawyer, we can find that she has taken charge of some cases, where each case is corresponding to a textual description and is related to several fields of crimes (i.e., "*Murder*" and "*Affray*" in "Case 1"). Generally speaking, before the trial, the lawyer meet with the client to discuss the anticipatory strategies to be pursued at the trial from relevant evidence and legal articles. In court, the trial lawyer should present evidence and claims and properly refute the evidence and words that against the client with adversarial lawyers (i.e., plaintiff or defendant) during the trial [11]. After a series of debates, they receive a final court judgment, from which the effects of lawyers, in this case, can be evaluated (e.g., in "Case 2", the lawyer argued that the burglar had voluntarily confessed to the crime and that it was the first time, and the judge accepted the opinion and reduced the punishment. We believe the lawyer had a positive effectiveness on this case while maintaining justice). Naturally, through analyzing "Case 1", we can roughly assess her skills in "*Murder*" and "*Affray*" fields. Through all of her careers, we can get a comprehensive assessment of her fields of expertise (i.e., *Murder*, *Affray*, *Steal*, *Fraud* and *Illegally Storing Guns*, as shown in the radar map of Figure 1).

In the literature, there are many studies on Legal Intelligence by means of large number of unstructured textual resources in this domain. Some studies focus on legal judgment prediction and

[1]https://www.findlaw.com, https://www.51djl.com/

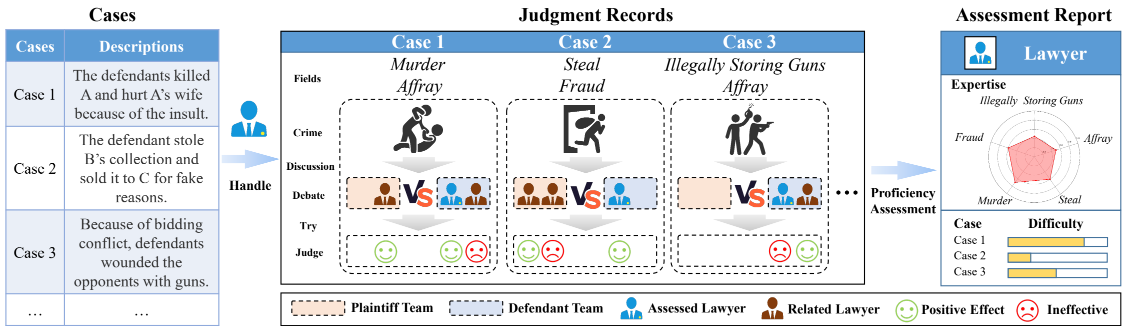


Figure 1: An example of lawyer assessment. The lawyer handles 3 cases and leaves the judgment records related to 5 fields. Specifically, the judgment records show the process and details of litigations. With proficiency assessment methods, we get the assessment report containing the lawyer's proficiency in each field and difficulties of each case.

information retrieval using judgment documents [21, 29, 37, 39]. The others are about legal question answering according to the legal articles and court records [7, 40]. Unfortunately, few studies have been done from the perspective of lawyer profiling except some social network-based approaches [3, 26]. To this end, driven by the actual needs of this legal domain, we propose a focused study on assessing proficiency of trial lawyers.

As a matter of fact, many technological and domain challenges are inherent in designing an automatic solution to this problem. In the first place, it is unavoidable to rely on judgment results of cases when assessing the proficiency of trial lawyers [28, 41]. In reality, different lawyers might show various proficiencies in the same case, and we still lack a well-defined quantifiable standard of expertise. Each lawyer is usually responsible for some specific fields. Thus, it is unfair to compare them directly. Things get more complicated when it comes to lawyers with different fields of expertise in charge of different cases. Second, the litigation is not only about one lawyer, but also a debate involving other lawyers and adversarial circumstance, meaning that we have to start from multiple perspectives (i.e., not only from a individual perspective but also from an adversarial perspective) to assess the target lawyer. Third, although the proficiency of lawyers has a significant influence on the final judgment of cases [15, 31], the characteristics of the case itself also play an essential role (e.g., the difficulty and discrimination of cases). It is obvious that a more challenging and unusual lawsuit puts a higher demand on proficiencies of lawyers.

To conquer these challenges, in this paper, we first propose a Lawyer Proficiency Assessment Network (LawyerPAN) for mining proficiency of lawyers in various fields through litigation records of lawyers. Specifically, as we cannot obtain the true field proficiency of lawyers in the real world, we leverage lawyer effectiveness prediction task to automatically learn the lawyer proficiency. First, LawyerPAN maps the lawyer, case and team factors in each field into a unified real number space ranging from 0 to 1. Since the importance of cases factors, we quantify the case difficulty by encoding textual descriptions of cases, and characterize the case discrimination with the maximum difference of the proficiencies of all lawyers in the plaintiff or defendant team. Then, considering the legal scene's complexity, LawyerPAN models the interactions between lawyers and cases from multiple perspectives (i.e., the anticipatory and adversarial perspectives) with utilizing proposed joint probability formula for lawyer proficiency assessment. To be specific, the anticipatory view uses relative level of lawyer proficiency and case difficulty to indicate the probability that the lawyer might have a positive effectiveness on the case, while the adversarial view aims to depict the gap of lawyers' proficiency between both sides in the debate. Finally, we evaluate LawyerPAN by conducting extensive experiments on a real-world dataset [2], and the experimental results demonstrate the effectiveness and interpretability of LawyerPAN for assessing the lawyer proficiency in different fields. Significantly, proficiency assessment of trial lawyers can be of great help to match lawyers with cases in the corresponding fields and provide personalized legal service.

2 RELATED WORK

As far as we are concerned, few existing works have been directly designed for lawyer assessment, while much research about legal Intelligence can give us some inspirations. Besides, according to our investigation, there are potential similarities between lawyer profiling and some proficiency assessment approaches, so we have borrowed some ideas from these relevant research fields.

2.1 Legal Intelligence

Legal Intelligence mainly focuses on applying artificial intelligence technologies to traditional legal tasks. Generally, most data resources in this field are unstructured texts, such as judgment documents, contracts, and legal opinions. Therefore, most LegalAI tasks are based on Natural Language Processing technologies [17, 28, 37, 41] which could be mainly divided into three aspects. First, researchers[21, 39] have paid much attention to Legal Judgment Prediction (LJP), which automatically decides the judgment results based on case fact. For example, researchers in [39] explored the multiple subtasks of legal judgment and modeled the judgment process according to topological order of subtasks. Second, Legal Question Answering has been widely studied in Legal Intelligence, which aims at answering questions in the legal domain[7, 40]. For instance, researchers in [40] proposed a dataset collected from the bar exam and showed the performance of giving some baselines. Finally, Information Retrieval is a typical application in Legal Intelligence [29]. Researchers in [29] models the paragraph-level interactions of case documents for legal case retrieval task.

[2]https://www.51djl.com/

Especially in terms of lawyer profiling, few studies have been done from this perspective. Some studies are social network-based approaches, which focus on ranking lawyers and searching unusual events through legal cases [3, 26]. The others are based on social statistics [11, 27, 30]. Unfortunately, these methods show some limitations in exploiting cases to precisely represent the lawyers' proficiency level in each field because achieving this task has to not only handle multiple heterogeneous fields (e.g., *Steal*, *Fraud*) but also consider associations between the teammates and rivals.

2.2 Proficiency Assessment

Proficiency Assessment is a necessary and fundamental task in many real-world scenarios such as psychological education [35], games [4] and e-commerce [13].

Psychological education. In intelligent education systems [2, 19], cognitive diagnosis aims to discover the states of students in the learning process, such as their proficiencies on specific knowledge concepts [18]. DINA [5, 34] and IRT [10] are two of the most typical works from the educational psychology area, which model the result of a student answering an exercise as the interaction between the trait features of the student (θ) and the exercise (β). Specifically, in DINA, θ and β are binary, where β comes directly from Q-matrix (a human labeled exercise-knowledge correlation matrix). In IRT, θ and β are unidimensional and continuous latent traits, indicating student proficiency and exercise difficulty respectively. The interaction between the trait features is modeled in a logistic way (e.g., a simple version is $sigmoid(a(\theta - \beta))$), where a is the exercise discrimination parameter. Recently, deep learning models have also been applied successfully to diagnose students' cognition combined with traditional functions [35].

Games. Gaming research mostly emphasizes the tracking the pairwise influence and analyzes players or heroes, by assessing the synthetic ability score, or by exploring the high-order interactions among the heroes and teams [12]. This ability score is usually derived from the probabilistic algorithms like BradleyTerry, ELO, Glicko, TrueSkill [9, 14]. The ability of the team is then modeled as the summation of the team members' scores.

E-commerce. The factorization algorithm, which is one of the common methods in e-commerce, treats the recommendation as a user-item matrix reconstruction problem and model the user-item interactions by dot product of latent vectors [23, 36]. In recent years, deep learning has also been applied successfully to the classical collaborative filtering user-item matrix reconstruction problems from different perspectives [13].

However, due to the variable cooperation of lawyers and the complicated antagonistic relationship in litigation, the above evaluation methods are difficult to be directly applied to the assessment of lawyers' proficiency.

3 PRELIMINARIES

In order to better present the modeling process, we first introduce relevant psychological and educational theory (i.e., Item Response Theory [10, 20]). Then, we propose a joint probability formula of Lawyer Proficiency Assessment inspired by the theory of psychological education. Finally, we will give a formal problem formulation of proficiency assessment of lawyers.

3.1 Proficiency Assessment Modeling

3.1.1 ***Item Response Theory.*** IRT is one of the most important psychological and educational theories, which roots in psychological measurement [10, 20]. With the student latent trait θ, question discrimination a and difficulty β as parameters, IRT can predict the probability that the student answers a specific question correctly with the *Sigmoid* function, and the 2-parameter logistic model of IRT is formulated as:

$$P(\theta) = \frac{1}{1 + e^{-Da(\theta - \beta)}}, \tag{1}$$

where $P(\theta)$ is the correct probability, D is a constant which often set as 1.7 [10]. That is, the probability of correct response to the exercises is positively correlated with students' knowledge proficiency.

Though positive progress has been made by IRT in education intelligence, it lacks the capability of deep feature representations since only one dimension has been considered.

3.1.2 ***Lawyer Proficiency Assessment.*** Inspired by proficiency assessment approach and team ability modeling [12, 38], we formulate the joint probability for lawyer proficiency assessment as:

$$P(\theta) = \frac{1}{1 + e^{-Q \cdot (Max\Theta - Min\Theta)(\theta - \beta)}}. \tag{2}$$

Parameter Q ***and Q-matrix.*** Parameter Q represents Q-matrix [32], which is a human labeled binary correlation matrix which denotes that the corresponding relationship between cases and legal fields. $\boldsymbol{Q}_{i,j} = 1$ if case i relates to field j and $\boldsymbol{Q}_{i,j} = 0$ otherwise.

Parameter Θ***.*** Parameter Θ indicates all the proficiency of lawyers involved in this case. In most fields of law, lawyers need to master enough legal knowledge to be skilled in handling litigations. For each case, we regard the effectiveness of distinguishing lawyers' skills as **case discrimination**. Specifically, we utilize the difference between the most proficient and the least proficient lawyers to quantify the case discrimination.

Parameter θ ***and*** β***.*** Since strategy making and debate circumstance are involved in court trial proceedings, we use formula (2) to contrast two relationships: 1) Modeling the relative relationship between lawyer proficiency and case difficulty in anticipatory circumstance (i.e., θ stands for **lawyer proficiency** and β stands for **case difficulty**). That is, if the lawyer proficiency is higher than the case difficultly, the lawyer might have a positive effectiveness on this case. Further, the probability of positive effectiveness to the cases is positively correlated with lawyers' field proficiency [30]. 2) Comparing proficiencies of the two teams in adversarial circumstance (i.e., regard θ and β as **plaintiff and defendant team proficiency** respectively).

We will expand each parameter in the formula to multiple dimensions, and each dimension represents one field of expertise.

3.2 Proficiency Assessment Task

In this section, we define the input data and the modeling problem in this paper. We summarize the notations in Table 1. Specifically, we assume that there are N Lawyers, M Cases and K Fields at a legal system, which can be represented as $L = \{l_1, l_2, \ldots, l_N\}$, $C = \{c_1, c_2, \ldots, c_M\}$ and $F = \{f_1, f_2, \ldots, f_K\}$ respectively.

Table 1: Several important mathematical notations.

Notation	Description
$\boldsymbol{L}$	The set of lawyers.
$\boldsymbol{C}$	The set of cases.
$\boldsymbol{F}$	The set of fields.
$\boldsymbol{R}$	The set of judgment records.
$\boldsymbol{G}^*$	The set of lawyers in plaintiff or defendant team.
$\boldsymbol{h}^l$	The proficiency of lawyers.
$\boldsymbol{Q}_c$	The case-field mappings of cases.
$\boldsymbol{h}_*^{diff}$	The difficulty of cases when the lawyer agent for plaintiff or defendant.
$\boldsymbol{h}_*^{disc}$	The discrimination of cases.
$\boldsymbol{h}_*^{team}$	The team proficiency of lawyers in plaintiff or defendant team.
$\boldsymbol{T}^*$	The representation of plaintiff or defendant team.
$\gamma(\cdot)$	The uniformization function.
$\boldsymbol{x}^A$	The output of anticipatory module.
$\boldsymbol{x}^T$	The output of adversarial module.

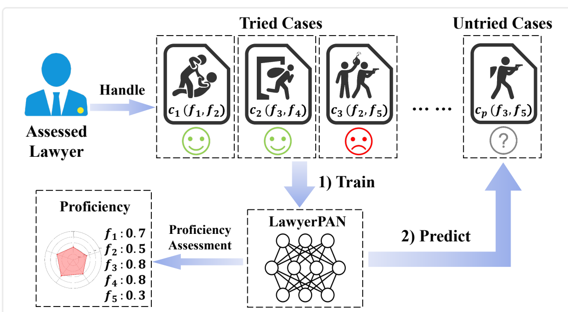


Figure 2: The framework of task.

Case logs. Each case includes lawyers in plaintiff and defendant teams denoted as $c_j = \left\{G_j^{pla}, G_j^{def}, D_j^w\right\}$, $G_j^* = \{l_1, l_2, \ldots, l_U\}$, which U represents the number of lawyers in corresponding team. Further, case descriptions are expressed as word sequences $D_j^w = \left\{d_{j,1}^w, d_{j,2}^w, \ldots, d_{j,n_w}^w\right\}$, where n_w denotes the length of the sequence. In addition, we have Q-matrix $\boldsymbol{Q} \in \mathbb{R}^{M\times K}$ to represent related fields mentioned in Section 3.1.2.

Lawyer records. Each lawyer handles some cases, and the judgment records are represented by R, which $R_{ij} = (l_i, c_j, y_{ij})$ denotes the lawyer l_i obtains judgment result y_{ij} on case c_j. The judgment result $y_{ij} = 1$ if lawyer l_i has positive effectiveness (e.g., compensation or reducing penalty) on case c_j and $y_{ij} = 0$ otherwise.

Definition 1 (*Problem Definition*). *Formally, given lawyers' judgment records R and the Q-matrix (i.e., $\boldsymbol{Q}$), our goal is to leverage the judgment records R to train a prediction model $\mathbb{M}$ (e.g., LawyerPAN), which can be used to estimate the proficiencies for lawyers on each field in the newly-conducted prediction (i.e., case c_p) through* ***Lawyer Effectiveness Prediction*** *process.*

As shown in Figure 2, our solution is a two-stage framework, which contains a training stage and a prediction stage: 1) In the training stage, given judgment records as well as contextual information of cases (i.e., lawyers in plaintiff and defendant team), we propose LawyerPAN to understand and represent all contextual information of each case c_i as corresponding predicted proficiencies h_i^l; 2) In the prediction stage, after obtaining the trained LawyerPAN, for each untried cases without judgment results, we could estimate its proficiency with the available contextual information.

Definition 2 (*Lawyer Effectiveness Prediction*). *Given lawyers L, handled cases C and the Q-matrix (i.e., $\boldsymbol{Q}$), the goal of lawyer effectiveness prediction task is to predict whether the assessed lawyers have positive effectiveness on the case.*

4 PROFICIENCY ASSESSMENT FOR LAWYERS

In this section, we describe the details of LawyerPAN, which is a proficiency assessment network. Figure 3 illustrates the structure of LawyerPAN. The entire architecture could be split into three parts in a high-level discussion. First, the initial input is an embedding layer which consists of lawyers, cases and team factors. Specifically, the fields, difficulties and discriminations of cases are taken into case factors. We use a normalized function to unify and represent team proficiencies. Second, assessment layer, which aims to generate relations between lawyers and cases in different fields, could automatically learn the proficiency of lawyers from anticipatory module and adversarial module. Finally, interaction layer and output layer are designed to learn the proficiency relations on different levels adaptively.

4.1 Input and Embedding Layer

Here we introduce representations of lawyer factors, case factors and, team factors in LawyerPAN.

4.1.1 ***Lawyer factors.*** In LawyerPAN, each lawyer is represented with a proficiency vector. The lawyer factor $\boldsymbol{h}^l$, which corresponds to the parameter $\boldsymbol{\theta}$ in formula (2), is obtained by multiplying the lawyer's one-hot representation $\boldsymbol{x}^l$. That is,

$$\boldsymbol{h}^l = \text{sigmoid}\left(\boldsymbol{x}^l \times \boldsymbol{E}^l\right), \tag{3}$$

where $\boldsymbol{h}^l \in (0,1)^{1\times K}$, $x^l \in \{0,1\}^{1\times N}$, trainable matrix $\boldsymbol{E}^l \in \mathbb{R}^{N\times K}$.

Specifically, the same trainable matrix $\boldsymbol{E}^l$ is used to learn the lawyer proficiency vector in the plaintiff or defendant team. In the same way, the input of team factors to LawyerPAN, which corresponds to the parameter $\boldsymbol{\Theta}$ in formula (2), is lawyers' one-hot representation of plaintiff and defendant team. Field proficiency of lawyers in plaintiff and defendant team are denoted as:

$$\boldsymbol{h}_{pla}^{team} = \text{sigmoid}\left(\boldsymbol{x}^p \times \boldsymbol{E}^l\right), \tag{4}$$

$$\boldsymbol{h}_{def}^{team} = \text{sigmoid}\left(\boldsymbol{x}^d \times \boldsymbol{E}^l\right), \tag{5}$$

in which $\boldsymbol{h}_{pla}^{team} \in (0,1)^{U_p\times K}$, $x^p \in \{0,1\}^{U_p\times N}$, $\boldsymbol{h}_{def}^{team} \in (0,1)^{U_d\times K}$, $x^d \in \{0,1\}^{U_d\times N}$. U_p and U_d are numbers of lawyers in the plaintiff and defendant team. x^p is the one-hot representation matrix of the plaintiff team. Similarly, x^d corresponds to defendant team.

4.1.2 ***Case factors.*** In reality, each case itself has objective facts, involving relevant legal principles and legal articles. We regard them as case factors. To make a general representation and more precise assessment, we adopt three case factors: field vector, difficulty vector and discrimination vector.

Field vector. For each case, the case factor $\boldsymbol{Q}_c$, which corresponds to the parameter $\boldsymbol{Q}$ in formula (2), is directly from the

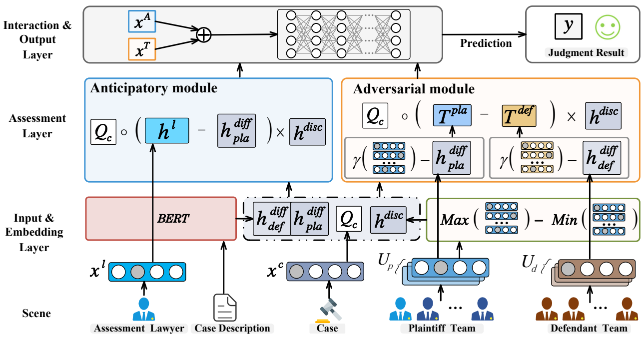


Figure 3: The illustration of the proposed LawyerPAN model. Overall, from the bottom up, it can be divided into 3 layers: 1) Input and Embedding Layer, which consists of assessment lawyers, cases and team factors; 2) Assessment Layer, which contains two modules to learn multiple complex relations between lawyers and cases; 3) Interaction and Output Layer, which is designed for feature learning and final prediction.

pre-given Q-matrix:

$$Q_c = x^c \times Q, \tag{6}$$

where $Q_c \in \{0,1\}^{1\times K}$, and $x^c \in \{0,1\}^{1\times M}$ is the one-hot representation vector of the case.

Difficulty vector. First, given a word sequence of the case description $D^w = \{d_1^w, d_2^w, \ldots, d_{n_w}^w\}$, we map each word of D^w into word embedding by BERT [6], and get the case embedding sequence $s^d \in \mathbb{R}^{d_c}$ by applying mean-pooling over x^w, where d_c is the dimension of case embedding, n_w is the length of word sequence. Then, $\boldsymbol{h}^{diff} \in (0,1)^{1\times K}$ indicates the difficulty of each field. We further use $\boldsymbol{h}_{pla}^{diff}$ and $\boldsymbol{h}_{def}^{diff}$ to distinguish difficulties of defending for the plaintiff and defendant which corresponds to the parameter $\boldsymbol{\beta}$ in formula (2). The difficulty factors can be obtained by:

$$\boldsymbol{h}_{pla}^{diff} = \text{sigmoid}\left(\boldsymbol{W}^p \boldsymbol{s}^d + \boldsymbol{b}^p\right), \tag{7}$$

$$\boldsymbol{h}_{def}^{diff} = \text{sigmoid}\left(\boldsymbol{W}^d \boldsymbol{s}^d + \boldsymbol{b}^d\right), \tag{8}$$

$$\boldsymbol{s}^d = MeanPool(BERT([d_1^w, d_2^w, \ldots, d_{n_w}^w])), \tag{9}$$

where $\boldsymbol{W}^*$ and $\boldsymbol{b}^*$ are learnable weight matrices and biases. Specifically, $\boldsymbol{W}^p$ and $\boldsymbol{W}^d$ show explainable relations of different difficulties between the plaintiff and defendant which lawyers handled.

Discrimination vector. $\boldsymbol{h}^{disc} \in (0,1)$ indicates the capability of the case to differentiate between those lawyers whose field proficiency is high from those with low field proficiency. The discrimination factors can be obtained by:

$$h_{pla}^{disc} = \text{sigmoid}\left(\boldsymbol{W}^c(Max(\boldsymbol{h}_{pla}^{team}) - Min(\boldsymbol{h}_{pla}^{team})) + \boldsymbol{b}^c\right), \tag{10}$$

$$h_{def}^{disc} = \text{sigmoid}\left(\boldsymbol{W}^c(Max(\boldsymbol{h}_{def}^{team}) - Min(\boldsymbol{h}_{def}^{team})) + \boldsymbol{b}^c\right). \tag{11}$$

4.1.3 ***Team factors.*** The compositions of plaintiff team and defendant team are essential for adversarial information of litigations. Intuitively, we first obtain the representations of all the lawyers in plaintiff team and defendant team, respectively. Then, we utilize a uniform function to represent the team proficiency of lawyers with lower dimensions. Finally, we utilize the team factors and the case factors to express the gap of team proficiency and case difficulty in plaintiff and defendant teams.

After obtaining representations of all the lawyers in plaintiff and defendant team (i.e., $\boldsymbol{h}_{pla}^{team}$ and $\boldsymbol{h}_{def}^{team}$) in formula (4) and (5), we indicate the team representations with the gap of team proficiency and case difficulty of both teams as:

$$\boldsymbol{T}^{pla} = \gamma\left(\boldsymbol{h}_{pla}^{team}\right) - \boldsymbol{h}_{pla}^{diff}, \tag{12}$$

$$\boldsymbol{T}^{def} = \gamma\left(\boldsymbol{h}_{def}^{team}\right) - \boldsymbol{h}_{def}^{diff}, \tag{13}$$

where $\gamma(\cdot)$ is a function which is proposed to uniformization the U-dimensional representation vector of the team. Intuitively, the overall level of the team depends more on the lawyer with the highest proficiency. Therefore, we formulate it as:

$$\gamma(\boldsymbol{V}) = \sum_{i=0}^{U} \text{softmax}(\boldsymbol{V})_i \circ \boldsymbol{V}_i, \tag{14}$$

where $\boldsymbol{V} \in (0,1)^{U\times K}$, $\boldsymbol{\gamma}(V) \in (0,1)^{1\times K}$.

4.2 Assessment Layer

For a more detailed assessment of the lawyer's proficiency in each field, we model the relationship between lawyers and cases in two scenarios mentioned in Section 3.1.2. That is anticipatory module to model strategy making proceedings, and adversarial module to model the debate in court trial proceedings.

4.2.1 ***Anticipatory module.*** To exploit relationships between assessed lawyers and case difficulties, the anticipatory module denotes the probability of higher proficiency compared with the case in the covered field. Following by formula (2), we formulate it as:

$$\boldsymbol{x}^A = \begin{cases} \boldsymbol{Q}_c \circ \left(\boldsymbol{h}^l - \boldsymbol{h}_{pla}^{diff}\right) \times \boldsymbol{h}_{pla}^{disc} & l \in G^{pla} \\ \boldsymbol{Q}_c \circ \left(\boldsymbol{h}^l - \boldsymbol{h}_{def}^{diff}\right) \times \boldsymbol{h}_{def}^{disc} & l \in G^{def} \end{cases}, \tag{15}$$

where $\circ$ is an element-wise product, l indicates the assessed lawyer, G^{pla} and G^{def} represents all of the lawyers which handled corresponding cases in plaintiff and defendant team.

4.2.2 ***Adversarial module.*** In this module, to probe relationships between lawyers and cases in plaintiff and defendant teams, we formulate the probability of relative superiority for lawyers between plaintiff and defendant teams as:

$$\boldsymbol{x}^T = \begin{cases} \boldsymbol{Q}_c \circ \left(\boldsymbol{T}^{pla} - \boldsymbol{T}^{def}\right) \times \boldsymbol{h}_{pla}^{disc} & l \in G^{pla} \\ \boldsymbol{Q}_c \circ \left(\boldsymbol{T}^{def} - \boldsymbol{T}^{pla}\right) \times \boldsymbol{h}_{def}^{disc} & l \in G^{def} \end{cases}. \tag{16}$$

4.3 Interaction and Output Layer

To adaptively integrate proficiency information, we concatenate representations of anticipatory and adversarial module. Followings are full connection layers and a prediction layer:

$$\begin{aligned} \boldsymbol{o}_1 &= tanh\left(\boldsymbol{W}_1 \,\text{Concat}\left(\boldsymbol{x}^A, \boldsymbol{x}^T\right)\right), \\ \boldsymbol{o}_2 &= tanh\left(\boldsymbol{W}_2 \boldsymbol{o}_1\right), \\ \boldsymbol{y} &= sigmoid\left(\boldsymbol{W}_3 \boldsymbol{o}_2\right). \end{aligned} \tag{17}$$

To satisfy the positively correlation between the probability $\boldsymbol{x}^A$, $\boldsymbol{x}^T$ and prediction result $\boldsymbol{y}$ mentioned in Section 3.1.2, we adopt a strategy which restricts each weights of fully connection layer (e.g., $\mathbf{W}_1$, $\mathbf{W}_2$ and $\mathbf{W}_3$) to be positive [35]. It can be easily proved that $\frac{\partial y}{\partial h_i^l}$ is positive for each entry h_i^l in $\boldsymbol{h}^l$. Thus the positively correlation is always satisfied during training.

4.4 Learning

The loss function of LawyerPAN is cross entropy between the ground truth y and prediction probability of $\hat{y}$:

$$\mathcal{L} = -\sum_i \left(y_i \log \hat{y}_i + (1-y_i)\log(1-\hat{y}_i)\right). \tag{18}$$

After training, the value of $\boldsymbol{h}^l \in (0,1)^{1\times K}$ is the assessment result, which denotes the lawyer's proficiency in K fields.

5 EXPERIMENTS

In this section, we demonstrate the effectiveness of LawyerPAN. First, we compare our LawyerPAN with some baselines and its variants on real-world datasets. Then, we make some interpretation assessments of models. As we cannot obtain the true field proficiency of lawyers in the real world, the assessment results are usually acquired through predicting lawyers' effectiveness, which can indirectly evaluate the model. Finally, case studies from individual and team perspectives are used to demonstrate the effectiveness and interpretability of lawyer assessments in the real world.

5.1 Dataset Description

As there are no publicly available datasets for Lawyer Proficiency Assessment, we construct LCHR [3], a *L*awyer *C*ases *H*andling *R*ecords dataset, which is collected from the 51djl.com in China. Specifically, LCHR mainly contains lawyers, cases, legal fields and judgment records. Among them, case logs consist of fields, case descriptions, and participating lawyers. Many cases contain multiple plaintiffs and defendants. The cases are settled from 2014 to 2017 in a certain region of China, where all the dismissed cases and executive cases have been removed. Lawyer records are consist of judgment records (i.e., a set of case logs which the lawyer handled) and judgment results (i.e., labels of whether the lawyer has positive effectiveness on the case). Besides, the labels of the fields (i.e., Q-matrix) and judgment results are summarized from the judgment documents and labeled by legal experts who have no conflict of interest with those involved in these cases.

In the data processing, for effective training, we construct LCHR-small by deleting cases which contain only one lawyer on the basis of LCHR. As for lawyer records, we filter out lawyers with less than 10 judgment records respectively to guarantee that each lawyer has enough data for assessment. Meanwhile, we divide all datasets into training, validation, and test sets according to the ratio of 7:1:2. To be consistent with the reality (i.e., any case will be handled only once), we remove the validation and test cases which appear in the training set. Table 2 summarizes basic statistics of the whole datasets. In addition, some fine-grained statistics of LCHR-small are shown in Figure 4. Specifically, the left part is a histogram chart denoting the distribution of the effectiveness of lawyers, where we can find the tremendously different levels of the lawyers' proficiency. The right part shows the effectiveness of a certain civil majored lawyer for different legal fields (e.g., *civil dispute, contract dispute*), which supports the idea that the proficiency varies greatly for different legal fields of the one lawyer.

[3] https://github.com/bigdata-ustc/LawyerPAN

Table 2: Dataset Statistics.

Dataset	LCHR-small	LCHR
# Lawyers	26,430	29,015
# Cases	193,444	77,2920
# Fields	525	636
# Judgment records	440,614	1,020,090
# Lawyers per case	2.442	1.623
# Causes per case	1.005	1.004

(a) Overall Distribution of Lawyers (b) Individual Distribution of One Lawyer

Figure 4: The distribution statistics of lawyer effectiveness in LCHR-small dataset.

5.2 Experimental Setup

Parameters setting. As the descriptions of facts are written by Chinese court clerk, word embeddings are obtained by the pre-trained "BERT-base, Chinese" model of BERT [6]. We adopt the second-to-last layer of BERT to train our model, and all the parameters of BERT model are set to the frozen state. Meanwhile, we set the maximum document length as 510 words. For training, the learning rate of Adam optimizer [16] is initialized as 10^{-4}. We train every model for 20 epochs with batch size 64 and choose the best model on the validation set for testing. To evaluate the performance of our model, all baseline models are implemented by Python, and all experiments are run on a Linux server with four 2.30GHz Intel Xeon Gold 5218 CPUs and a Tesla V100 GPU.

Evaluation metrics. The performance of a proficiency assessment model is difficult to evaluate as we cannot obtain the true legal field proficiency of lawyers in the real world. As the assessment results are usually acquired through predicting lawyers' effectiveness, performance on this prediction task can indirectly evaluate the model. Therefore, for predicting the positive effectiveness of lawyers, we choose four evaluation metrics from both classification aspect and regression aspect, including Accuracy, RMSE (Root Mean Square Error), AUC (Area Under the Curve), and F1-score. Generally speaking, the better prediction has a high value among Accuracy, AUC, and F1-score. And the lower values of the RMSE values are, the better prediction is obtained.

5.3 Baseline Approaches

To evaluate the performance of our model on lawyer effectiveness prediction task, we selecte a number of representative methods as baselines. Specifically, we first choose some general and basic approaches as follows:

- **LR** [24]: a linear model with logistic loss for classification.
- **AVG**: a statistical-based approach by taking an average of the effectiveness value in all cases of each lawyer.

Table 3: The prediction performance of lawyers' effectiveness on cases. The values in brackets of LawyerPAN indicate the relative increase over NeuralCDM-BERT.

Methods	LCHR-small				LCHR			
	Accuracy	RMSE	AUC	F1	Accuracy	RMSE	AUC	F1
LR	0.709	0.455	0.649	0.829	0.694	0.458	0.652	0.819
AVG	0.706	0.443	0.663	0.810	0.715	0.438	0.685	0.816
BERT	0.723	0.416	0.763	0.805	0.731	0.412	0.779	0.804
PMF	0.504	0.459	0.500	0.588	0.498	0.462	0.494	0.579
IRT	0.701	0.484	0.523	0.823	0.693	0.506	0.498	0.818
TrueSkill	0.700	0.483	0.663	0.824	0.693	0.447	0.681	0.819
NeuralCDM	0.720	0.423	0.749	0.808	0.756	0.412	0.785	0.844
NeuralCDM-BERT	0.749	0.399	0.799	0.829	0.756	0.396	0.815	0.828
LawyerPAN	**0.878** (+0.13)	**0.303** (-0.10)	**0.929** (+0.13)	**0.914** (+0.09)	**0.885** (+0.13)	**0.291** (-0.11)	**0.941** (+0.13)	**0.917** (+0.09)
LawyerPAN-EI	0.873	0.304	0.931	0.910	0.882	0.294	0.940	0.915
LawyerPAN-SC	0.765	0.390	0.826	0.838	0.810	0.362	0.868	0.866
LawyerPAN-AM	0.877	0.306	0.926	0.912	0.879	0.296	0.937	0.913

- **BERT** [6]: a word encoder representation model using one-hot vectors of lawyers and case descriptions.

Then, we borrow some baselines from various domains of recommender system, psychological measurement and education:

- **PMF** [23]: a latent factor model which projects the users and items into a low-dimensional space by mining the sparse consumption matrix.
- **IRT** [1, 25]: a cognitive diagnosis method, which models students' latent traits and the parameters of exercises like difficulty and discrimination with a logistic-like function.
- **TrueSkill** [14]: a rating system among game players. It works well for both teams with different numbers of players.
- **NeuralCDM** [35]: a neural cognitive diagnostic framework, which leverages multi-layers for modeling the complex interactions of students and exercises, aims to diagnose students' cognition by predicting the probability of the student answering the exercise correctly.
- **NeuralCDM-BERT**: a NeuralCD model where exercise difficulty was replaced by BERT while keeping the dimension of exercise difficulty representation unchanged.

Particularly, when adopting the above models in lawyer effectiveness prediction, we replace relevant concepts (i.e., students and exercises, users and items, both players) as lawyers and cases.

Finally, to further validate the performance of each component in our model, we also design some simplified variants, including:

- **LawyerPAN-EI**: it replaces the case discrimination with the case one-hot embedding vector.
- **LawyerPAN-AM**: it only adopts the anticipatory module in assessment layer.
- **LawyerPAN-SC**: it uses the same case factors on the defendant and plaintiff (i.e., $\boldsymbol{h}^{diff}_{pla} = \boldsymbol{h}^{diff}_{def}, \boldsymbol{h}^{disc}_{pla} = \boldsymbol{h}^{disc}_{def}$).

5.4 Experimental Results

To evaluate the overall performance of LawyerPAN on lawyer effectiveness prediction task, we first show the comparison results of LawyerPAN and its variants with baselines in Table 3. Then, we compare several uniformization functions to verify the rationality of team proficiency representation. Finally, the visualization of assessment parameters demonstrated the reasonable relationship between learned lawyer proficiency and case difficulty.

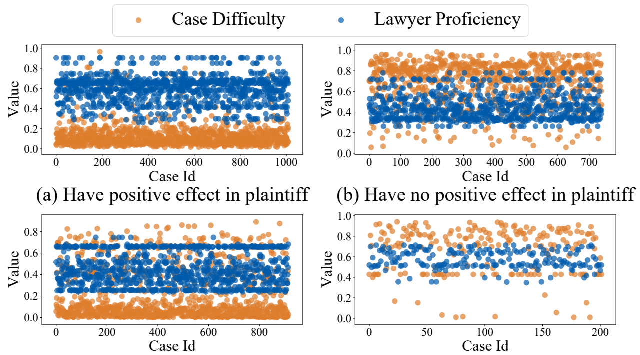
Table 4: Experimental results on uniformization function.

Methods	Accuracy	RMSE	AUC	F1
# MeanPooling	0.876	0.306	0.926	0.912
# MaxPooling	0.873	0.308	0.926	0.910
# MinPooling	0.876	0.308	0.926	0.912
# Softmax (LawyerPAN)	**0.878**	**0.303**	**0.929**	**0.914**

5.4.1 ***Overall performance.*** Obviously, our LawyerPAN and its variants achieve the better performance than other baselines observing from Table 3. Then, comparing with other baselines, we have the following observations from the results. 1) We compare basis approaches (i.e., LR, AVG and BERT) with our model, which all have poor performance because the LawyerPAN exploits deep interactions between cases and lawyers. 2) Contrasted to some baselines from different domains, we can find that our model, which have the clear correspondence between lawyer latent features and legal fields, performs better than PMF, IRT and TrueSkill. 3) NeuralCDM and NeuralCDM-BERT outperform other baselines but our LawyerPAN and its variants exceed them, which indicates that the LawyerPAN can model the relationship between the lawyer and corresponding cases effectively while ignoring the adversarial relevance of lawyers' compositions.

5.4.2 ***Ablation study.*** We focus on LawyerPAN and its variants. Particularly, for LawyerPAN-EI, it adopts the case one-hot embedding vector to represent the case discrimination. It performs poorer than LawyerPAN, which not only demonstrates the importance of the difference between the most proficient and the least proficient lawyer, but also shows the effectiveness of our proposed joint probability mentioned in Section 3.1.2. Next, we adopt LawyerPAN-SC by assuming that the difficulty and discrimination of cases are the same for all litigants. From the observation which LawyerPAN performs better than LawyerPAN-SC, we can demonstrate that the effectiveness of designing different case difficulty for corresponding litigants in LawyerPAN. Finally, the only difference between the variants, LawyerPAN-AM and NeuralCDM-BERT, lies in utilizing probability formulas (1) and (2) respectively. The more precise performance of LawyerPAN-AM proves the effectiveness of joint probability formula (2).

5.4.3 ***Impact of uniformization function.*** Afterwards, due to the impact of the adversarial module on proficiency assessment, we turn to discuss this adversarial module. Specifically, in the practical



(a) Have positive effect in plaintiff (b) Have no positive effect in plaintiff

(c) Have positive effect in defendant (d) Have no positive effect in defendant

Figure 5: Visualization of assessment parameters.

litigation context, debates in court are usually conducted in teams. Intuitively, the comprehensive proficiency of the lawyer teams depends on each lawyer involved in the litigations. Therefore, we make a series of experiments about the variants of formula (14) to discuss this situation in Table 4. Among them, mean-pooling, max-pooling and min-pooling represent the average, maximum and minimum proficiency assessment respectively. From the observation, we can find that the mean-pooling approach has a little better result than max-pooling and min-pooling, which demonstrates the importance of overall proficiency of lawyers. Meanwhile, different from the above methods, the formula (14) we designed considers the impact of each lawyer. Further, the better performance of proposed uniformization function shows the lawyers with higher proficiency play a more important role in the litigation process. Therefore, at the same time of getting the best result, it is also consistent with our intuitive understanding.

5.4.4 ***Visualization of assessment parameters.*** To compare with the lawyer proficiency, our LawyerPAN model evaluate the difficulty of the cases and Figure 5 shows the value of lawyer proficiency and case difficulty in the field "*housing presale contract dispute*". We assume that if the lawyer proficiency is higher than the case difficultly, the lawyer might have a positive effectiveness on this case mentioned in Section 3.1.2. Through the observation, we find that, in most cases where lawyers have a positive effectiveness in the plaintiff and defendant, lawyer proficiency value (i.e., blue points) is greater than case difficulty value (i.e., yellow points). In the cases where lawyers have no positive effectiveness, the difficulty value of most cases is greater. The model evaluation results are consistent with our hypothesis, and this further suggests that there is a huge distinction between a lawyer agent for the plaintiff and defendant.

5.5 Case Study

To demonstrate the effectiveness and interpretability of the assessment in the real world, we provide a qualitative analysis of LawyerPAN from individual and team perspectives.

5.5.1 ***Individual perspective.*** We give an example of the assessment reports containing the individual lawyers' proficiency in each legal field in Figure 6. Here, we demonstrate the rationality and effectiveness of LawyerPAN from two aspects. On the one hand, in the field f_1 (i.e., *Sales Contract Dispute*), since both lawyers have

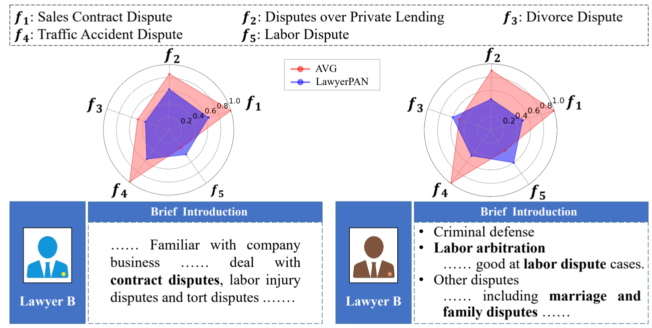


Figure 6: An assessment example of 2 lawyers. The radar figure shows proficiency assessment on 5 common fields with similar average effectiveness. Bottom descriptions are excerpted from the resumes on their home page of websites.

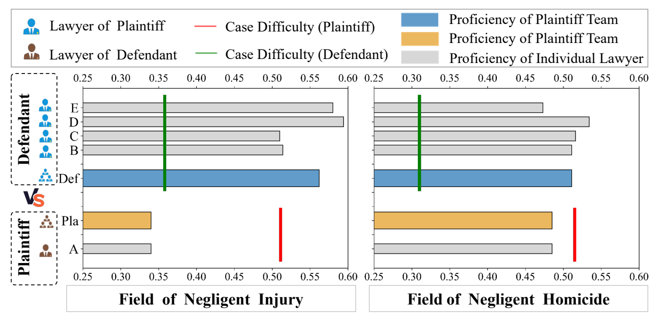


Figure 7: A case study of lawyer assessment in the adversarial module of LawyerPAN.

positive effectiveness on all handling records, the evaluation result of AVG method is 1.0, which is unlikely to reach in reality. LawyerPAN apparently has more objective assessment results. On the other hand, generally, due to the different difficulty of the cases and the different record of handling cases, the same proficiency values would be obtained difficultly. For instance, lawyer B is more proficient than lawyer A by assessing with LawyerPAN in field f_3 (i.e., *Divorce Dispute*), which is consistent with the description of "marriage and family disputes" in the resume of the lawyer B. In this sense, the LawyerPAN is in line with the reality.

5.5.2 ***Team perspective.*** Furthermore, to intuitively verify this interpretability, we design a visualization in Figure 7 on the perspective of adversarial teams. This example shows a criminal case concerning two legal fields (i.e., *negligent injury* and *homicide*). There is one lawyer in plaintiff team and four lawyers in defendant team. The yellow and blue bar represent the comprehensive proficiency of plaintiff and defendant team respectively, and the grey shows the personal proficiency. The different lines denote different case difficulties. Specifically, the case difficulty is higher than the proficiency of the lawyer in plaintiff with both fields, which shows the lawyer in plaintiff team hardly has positive effectiveness on this case. Meanwhile, for the whole team in defendant, the case difficulty (i.e. green line) is lower than lawyers' comprehensive proficiency (i.e. blue bar). It reflects defendant team can handle this case effectively. From the above observations, we can conclude that our LawyerPAN can give interpretable results by exploiting both case difficulty and lawyer proficiency.

6 CONCLUSION

In this paper, we presented a focused study on proficiency assessment for trial lawyers and proposed a LawyerPAN model for mining proficiency of lawyers in various fields. To be specific, we designed the anticipatory module to exploit the lawyer's proficiency of anticipatory solution and strategy making in a certain case, and the adversarial module to depict the gap of lawyers' proficiency between both sides in the debate. Finally, we evaluated our approaches by conducting extensive experiments on our collected data, and the experimental results clearly demonstrated the effectiveness and interpretability of our proposed approaches. In summary, Lawyer Proficiency Assessment can help people who are unfamiliar with legal expertise and lawyers to choose appropriate legal teams when they have trouble in lawsuits. Therefore, we hope this work can help more people and boost much research in this promising field.

ACKNOWLEDGMENTS. This research was partially supported by grants from the National Key Research and Development Program of China (No. 2018YFC0832101), and the National Natural Science Foundation of China (Grant No.61922073).

REFERENCES

[1] A Lord Birnbaum. 1968. Some latent trait models and their use in inferring an examinee's ability. *Statistical theories of mental test scores* (1968).
[2] Hugh Burns, Carol A Luckhardt, James W Parlett, and Carol L Redfield. 2014. *Intelligent tutoring systems: Evolutions in design.* Psychology Press.
[3] Angelo Mondaini Calvão, Crysttian Arantes Paixão, Flávio Codeco Coelho, and Renato Rocha Souza. 2015. The consumer litigation industry: Chasing dragon kings in lawyer–client networks. *Social Networks* 40 (2015), 17–24.
[4] Shuo Chen and Thorsten Joachims. 2016. Predicting matchups and preferences in context. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.* 775–784.
[5] Jimmy De La Torre. 2009. DINA model and parameter estimation: A didactic. *Journal of educational and behavioral statistics* 34, 1 (2009), 115–130.
[6] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. [n.d.]. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics.* 4171–4186.
[7] Xingyi Duan, Baoxin Wang, Ziyue Wang, Wentao Ma, Yiming Cui, Dayong Wu, Shijin Wang, Ting Liu, Tianxiang Huo, Zhen Hu, et al. 2019. Cjrc: A reliable human-annotated benchmark dataset for chinese judicial reading comprehension. In *China National Conference on Chinese Computational Linguistics.* Springer, 439–451.
[8] NA Elbers, AJ Akkermans, P Cuijpers, DJ Bruinvels, et al. 2012. Exploring Lawyer-Client Interaction: A Qualitative Study of Positive Lawyer Characteristics. *Psychological injury and law* 5, 1 (2012), 89–94.
[9] Arpad E Elo. 1978. *The rating of chessplayers, past and present.* Arco Pub.
[10] Susan E Embretson and Steven P Reise. 2013. *Item response theory.* Psychology Press.
[11] Monroe H Freedman. 1966. Professional responsibility of the criminal defense lawyer: The three hardest questions. *Michigan Law Review* 64, 8 (1966), 1469–1484.
[12] Linxia Gong, Xiaochuan Feng, Dezhi Ye, Hao Li, Runze Wu, Jianrong Tao, Changjie Fan, and Peng Cui. 2020. OptMatch: Optimized Matchmaking via Modeling the High-Order Interactions on the Arena. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining.* 2300–2310.
[13] Xiangnan He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, and Tat-Seng Chua. 2017. Neural collaborative filtering. In *Proceedings of the 26th international conference on world wide web.* 173–182.
[14] Ralf Herbrich, Tom Minka, and Thore Graepel. 2006. TrueSkill™: a Bayesian skill rating system. In *Proceedings of the 19th international conference on neural information processing systems.* 569–576.
[15] Christel Karsten, Ulrike Malmendier, Zacharias Sautner, et al. 2014. *M&A negotiations and lawyer expertise.* Technical Report. Working Paper.
[16] Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. In *Proceedings of International Conference on Learning Representations.*
[17] Qi Liu, Han Wu, Yuyang Ye, Hongke Zhao, Chuanren Liu, and Dongfang Du. 2018. Patent Litigation Prediction: A Convolutional Tensor Factorization Approach.. In *Proceedings of the International Joint Conference on Artificial Intelligence.* 5052–5059.
[18] Qi Liu, Runze Wu, Enhong Chen, Guandong Xu, Yu Su, Zhigang Chen, and Guoping Hu. 2018. Fuzzy cognitive diagnosis for modelling examinee performance. *ACM Transactions on Intelligent Systems and Technology (TIST)* 9, 4 (2018), 1–26.
[19] Zitao Liu, Guowei Xu, Tianqiao Liu, Weiping Fu, Yubi Qi, Wenbiao Ding, Yujia Song, Chaoyou Guo, Cong Kong, Songfan Yang, et al. 2020. Dolphin: A spoken language proficiency assessment system for elementary education. In *Proceedings of The Web Conference 2020.* 2641–2647.
[20] Frederic M Lord. 1980. *Applications of item response theory to practical testing problems.* Routledge.
[21] Bingfeng Luo, Yansong Feng, Jianbo Xu, Xiang Zhang, and Dongyan Zhao. 2017. Learning to Predict Charges for Criminal Cases with Legal Basis. In *Proceedings of Empirical Methods in Natural Language Processing.* 2727–2736.
[22] Craig A McEwen, Nancy H Rogers, and Richard J Maiman. 1994. Bring in the lawyers: Challenging the dominant approaches to ensuring fairness in divorce mediation. *Minn. L. Rev.* 79 (1994), 1317.
[23] Andriy Mnih and Russ R Salakhutdinov. 2008. Probabilistic matrix factorization. In *Advances in neural information processing systems.* 1257–1264.
[24] Andrew Y Ng and Michael I Jordan. 2002. On discriminative vs. generative classifiers: A comparison of logistic regression and naive bayes. In *Advances in neural information processing systems.* 841–848.
[25] Georg Rasch. 1961. On general laws and the meaning of measurement in psychology. In *Proceedings of the fourth Berkeley symposium on mathematical statistics and probability*, Vol. 4. 321–333.
[26] Leonardo Filipe Rodrigues Ribeiro and Daniel Ratton Figueiredo. 2017. Ranking lawyers using a social network induced by legal cases. *Journal of the Brazilian Computer Society* 23, 1 (2017), 6.
[27] Rebecca L Sandefur. 2015. Elements of professional expertise: Understanding relational and substantive expertise through lawyers' impact. *American Sociological Review* 80, 5 (2015), 909–933.
[28] Tara Prasad Sapkota, Shila Kunwar, Mahima Bhattarai, and Shreya Poudel. 2020. Artificial intelligence that are beneficial for law. *US-China Law Review* 17, 5 (2020), 217–223.
[29] Yunqiu Shao, Jiaxin Mao, Yiqun Liu, Weizhi Ma, Ken Satoh, Min Zhang, and Shaoping Ma. 2020. BERT-PLI: Modeling Paragraph-Level Interactions for Legal Case Retrieval. In *Proceedings of the International Joint Conference on Artificial Intelligence.*
[30] Marjorie M Shultz and Sheldon Zedeck. 2011. Predicting lawyer effectiveness: Broadening the basis for law school admission decisions. *Law & Social Inquiry* 36, 3 (2011), 620–661.
[31] John Szmer, Susan W Johnson, and Tammy A Sarver. 2007. Does the lawyer matter? Influencing outcomes on the Supreme Court of Canada. *Law & Society Review* 41, 2 (2007), 279–304.
[32] Kikumi K Tatsuoka. 1984. Analysis of Errors in Fraction Addition and Subtraction Problems. Final Report. (1984).
[33] Bruce Tonn, Dorian Stiefel, John M Scheb II, Colin Glennon, and Hemant Kumar Sharma. 2012. Future of the courts: Fixed, flexible, and improvisational frameworks. *Futures* 44, 9 (2012), 802–811.
[34] Matthias von Davier. 2014. The DINA model as a constrained general diagnostic model: Two variants of a model equivalency. *Brit. J. Math. Statist. Psych.* 67, 1 (2014), 49–71.
[35] Fei Wang, Qi Liu, Enhong Chen, Zhenya Huang, Yuying Chen, Yu Yin, Zai Huang, and Shijin Wang. 2020. Neural Cognitive Diagnosis for Intelligent Education Systems. In *Proceedings of the AAAI Conference on Artificial Intelligence.*
[36] Le Wu, Junwei Li, Peijie Sun, Richang Hong, Yong Ge, and Meng Wang. 2020. DiffNet++: A Neural Influence and Interest Diffusion Network for Social Recommendation. *IEEE Transactions on Knowledge and Data Engineering* (2020).
[37] Linan Yue, Qi Liu, Binbin Jin, Han Wu, Kai Zhang, Yanqing An, Mingyue Cheng, Biao Yin, and Dayong Wu. 2021. NeurJudge: A Circumstance-aware Neural Framework for Legal Judgment Prediction. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval.*
[38] Hongke Zhao, Binbin Jin, Qi Liu, Yong Ge, Enhong Chen, Xi Zhang, and Tong Xu. 2019. Voice of charity: Prospecting the donation recurrence & donor retention in crowdfunding. *IEEE Transactions on Knowledge and Data Engineering* 32, 8 (2019), 1652–1665.
[39] Haoxi Zhong, Zhipeng Guo, Cunchao Tu, Chaojun Xiao, Zhiyuan Liu, and Maosong Sun. 2018. Legal Judgment Prediction via Topological Learning. In *Proceedings of Empirical Methods in Natural Language Processing.*
[40] Haoxi Zhong, Yuzhong Wang, Cunchao Tu, Tianyang Zhang, Zhiyuan Liu, and Maosong Sun. 2020. Iteratively questioning and answering for interpretable legal judgment prediction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 34. 1250–1257.
[41] Haoxi Zhong, Chaojun Xiao, Cunchao Tu, Tianyang Zhang, Zhiyuan Liu, and Maosong Sun. 2020. How Does NLP Benefit Legal System: A Summary of Legal Artificial Intelligence. *Proceedings of Annual Meeting of the Association for Computational Linguistics* (2020).