LawyerMTL: Lawyer Proficiency Assessment Network via Multi-task Learning

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.

Index Terms—Lawyer Proficiency Assessment, Legal Intelligence, Multi-Task Learning

1 Introduction

As an emerging trend, more and more legal platforms [1] come to provide online services [**?**] 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 [**?**].

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 "lose" the same way most non-professional people do [**?**], [**?**], [**?**]. 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 [**?**]. 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 focuse on legal judgment prediction and information retrieval using judgment documents [**?**], [**?**], [**?**], [**?**]. The others are about legal question answering according to the legal articles and court records [**?**], [**?**]. Unfortunately, few studies have been done from the perspective of lawyer profiling except some social network-based approaches [**?**], [**?**]. 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 [**?**], [**?**]. 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 [**?**], [**?**], 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

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

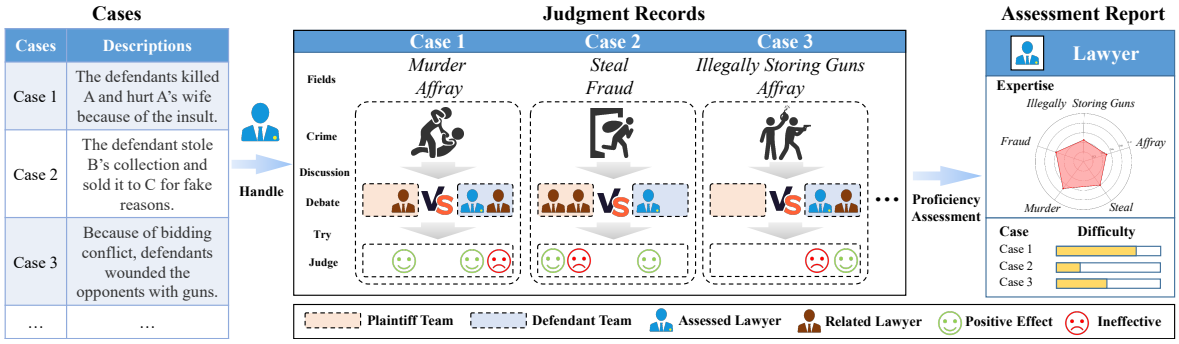


Fig. 1. An example of lawyer assessment. The lawyer handles 3 cases and leaves the judgment records related to 5 fields. Specifically, the judgment records show the process and details of litigations. With proficiency assessment methods, we get the assessment report containing the lawyer's proficiency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

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.

The purpose of our adoption of LawyerPAN for lawyer proficiency assessment is user personalization service oriented to legal intelligence, so we have made an improvement for the proficiency parameter, namely LawyerPANE. LawyerPANE adds constant initialization to ensure an initial proficiency parameter of 0.5. Moreover, the IRT parameters, proficiency parameters, and parameter fusion mode were explored and experimented in more depth in this paper.

To sum up, we made the following contributions in this paper:

- Inspired by [?], we propose a method to evaluate lawyers, which standardizes the ability of lawyers and the difficulty of cases into the same space. ¡br¿ We denote the competence of the plaintiff and the difficulty of the case separately.
- We propose the team factor, that is, the normalization function normalizes the ability of each lawyer in the team to the team proficiency.
- We use the proficiency difference between the lawyer plaintiff team and the defendant team to incorporate the lawyer effectiveness prediction.

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

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 [?], [?], [?], [?] which could be mainly divided into three aspects. First, researchers [?], [?] have paid much attention to Legal Judgment Prediction (LJP), which automatically decides the judgment results based on case fact. For example, researchers in [?] 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 [?], [?]. For instance, researchers in [?] 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 [?]. Researchers in [?] models the paragraph-level interactions of case documents for legal case retrieval task.

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 [?], [?].

The others are based on social statistics [?], [?], [?]. 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 [?], games [?] and e-commerce [?].

Psychological education. In intelligent education systems [?], [?], cognitive diagnosis aims to discover the states of students in the learning process, such as their proficiencies on specific knowledge concepts [?]. DINA [?], [?] and IRT [?] 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 [?].

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 [1]. This ability score is usually derived from the probabilistic algorithms like BradleyTerry, ELO, Glicko, TrueSkill [?], [?]. 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 [?], [?]. In recent years, deep learning has also been applied successfully to the classical collaborative filtering user-item matrix reconstruction problems from different perspectives [?].

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. There are two known effective methods for assessing a lawyer's proficiency: one is reviewing the lawyer's case handling record [2], and the other is evaluating their proficiency through conversation [3].

2.3 Multi-task Learning

2.4 Text Generation

[4] [5], [6], [7]

$$\mathcal{L} = \sum_{i=1}^{n} -y_i \log(\hat{y}_i) - (1-y_i)\log(1-\hat{y}_i) + \frac{1}{T}\sum_{t=1}^{T} -\log P(y_t^*) \tag{1}$$

3 Preliminaries

In order to better present the modeling process, we first introduce relevant psychological and educational theory (i.e., Item Response Theory [?], [?]). 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 [?], [?]. 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:

$$P(\theta) = \frac{1}{1 + e^{-Da(\theta-\beta)}}, \tag{2}$$

where $P(\theta)$ is the correct probability, D is a constant which often set as 1.7 [?]. 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 [?], [1], we formulate the joint probability for lawyer proficiency assessment as:

$$P(\theta) = \frac{1}{1 + e^{-Q\cdot(Max\Theta - Min\Theta)(\theta-\beta)}}. \tag{3}$$

Parameter Q and Q-matrix. Parameter Q represents Q-matrix [?], 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 (3) to contrast two relationships: 1) Modeling the relative relationship between lawyer proficiency and case difficulty in anticipatory circumstance (i.e., θ stands for

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.

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 [**?**]. 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.

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

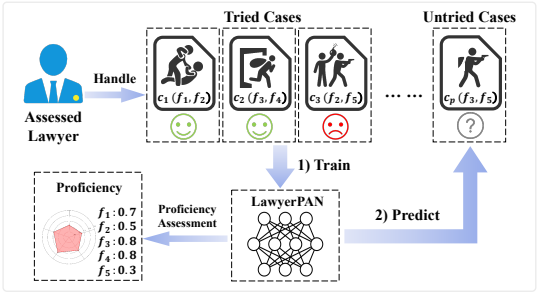


Fig. 2. The framework of task.

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, and the procedure is summarized in Algo. 1; 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, and the procedure is summarized in Algo. 2.

Algorithm 1 The LawyerPAN Training Framework

Input: lawyers L, handled cases C and the Q-matrix (i.e., $\boldsymbol{Q}$),$R = (l_i, c_j, r_{ij})$
Input: **Hyperparameters** $N_{epochs} \in \mathbb{N}$
Output: Trained model parameters h^l for downstream tasks

Initialize a neural network $\mathbb{M}$ f_θ
for each sampled judgment record R $\in Sampler(R)$ **do**
  for n=1...N **do**
    calculate $\hat{y} = \mathbb{M}(h^l)$
    $loss = -\sum_i \left(y_i \log \hat{y}_i + (1 - y_i) \log\left(1 - \hat{y}_i\right)\right).$
    $\theta = \theta - \eta.$
  end for
  Optimize θ by minimizing Loss function
  Update
end for
return h^l

Algorithm 2 The LawyerPAN Prediction Framework

Input: Trained model parameters h^l
$\hat{y} = \mathbb{M}(h^l)$
Output: Prediction result $\hat{y}$

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.*

In detail, as long as one of the claims of the client is satisfied (e.g., *compensation and reducing penalty*), we consider that the lawyers have positive effectiveness on the case. Generally, if the lawyer wins the lawsuit, he must have an influence on the case. Particularly, the lawyer may have an influence on the case if the lawyer loses the lawsuit.

The assumption is defined as follows: **Definition 3 (*Judgment Assumption*).** *In the case of absolute justice, the more proficient the lawyer, the greater the probability that the lawyer will have a 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 (3), is obtained by multiplying the lawyer's one-hot representation $\boldsymbol{x}^l$. That is,

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

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 (3), is lawyers' one-hot representation of plaintiff and defendant team. Field proficiency of lawyers in plaintiff and defendant team are denoted as:

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

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

in which $\boldsymbol{h}^{team}_{pla} \in (0,1)^{U_p\times K}$, $x^p \in \{0,1\}^{U_p\times N}$, $\boldsymbol{h}^{team}_{def} \in (0,1)^{U_d\times K}$, $x^d \in \{0,1\}^{U_d\times N}$. 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 (3), is directly from the pre-given Q-matrix:

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

where $\boldsymbol{Q}_c \in \{0,1\}^{1\times K}$, 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 $D^w = \{d^w_1, d^w_2, \ldots, d^w_{n_w}\}$, we map each word of D^w into word embedding by BERT [**?**], 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 $\boldsymbol{h}^{diff}_{pla}$ and $\boldsymbol{h}^{diff}_{def}$ to distinguish difficulties of defending for the plaintiff and defendant which corresponds to the parameter $\boldsymbol{\beta}$ in formula (3).

Since legal proceedings are different from other scenarios, the interaction records between lawyers and cases are sparse. If one-hot is used to represent the case, the effect is poor, so we adopt the text encoder to encode the case description text to represent the case. Since we use a text encoder to represent case difficulty rather than semantics, we believe that the choice of text encoder has little impact on the model effect. In addition, we changed the text encoder in LawyerPAN model from Bi-GRU to BERT because of the burst of BERT after the appearance. The difficulty factors can be obtained by:

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

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

$$\boldsymbol{s}^d = MeanPool(BERT([d^w_1, d^w_2, \ldots, d^w_{n_w}])), \tag{8}$$

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:

$$h^{disc}_{pla} = \text{sigmoid}\left(\boldsymbol{W}^c(Max(\boldsymbol{h}^{team}_{pla}) - Min(\boldsymbol{h}^{team}_{pla})) + \boldsymbol{b}^c\right), \tag{9a}$$

$$h^{disc}_{def} = \text{sigmoid}\left(\boldsymbol{W}^c(Max(\boldsymbol{h}^{team}_{def}) - Min(\boldsymbol{h}^{team}_{def})) + \boldsymbol{b}^c\right). \tag{9b}$$

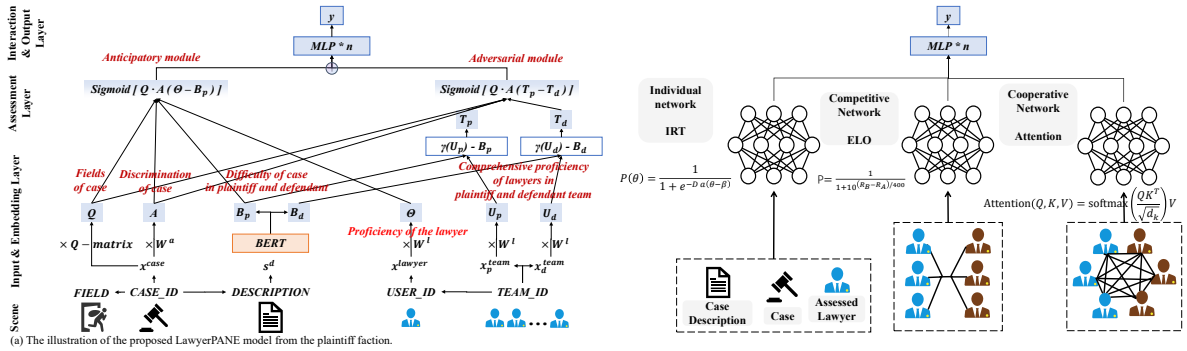
4.1.3 *Team factors.*

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., $\boldsymbol{h}^{team}_{pla}$ and $\boldsymbol{h}^{team}_{def}$) in formula (5a) and (5b), we indicate the team representations with the gap of team proficiency and case difficulty of both teams as:

$$\boldsymbol{T}^{pla} = \gamma\left(\boldsymbol{h}^{team}_{pla}\right) - \boldsymbol{h}^{diff}_{pla}, \tag{10a}$$

$$\boldsymbol{T}^{def} = \gamma\left(\boldsymbol{h}^{team}_{def}\right) - \boldsymbol{h}^{diff}_{def}, \tag{10b}$$



(a) The illustration of the proposed LawyerPANE model from the plaintiff faction.

Fig. 3. The illustration of the proposed LawyerPANE 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.

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:

$$\gamma(\boldsymbol{V}) = \sum_{i=0}^{U} \text{softmax}(\boldsymbol{V})_i \circ \boldsymbol{V}_i, \tag{11}$$

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 (3), we formulate it as:

$$\boldsymbol{x}^A = \begin{cases} \boldsymbol{Q}_c \circ \left(\boldsymbol{h}^l - \boldsymbol{h}_{pla}^{diff}\right) \times \boldsymbol{h}_{pla}^{disc} & l \in G^{pla} \\ \boldsymbol{Q}_c \circ \left(\boldsymbol{h}^l - \boldsymbol{h}_{def}^{diff}\right) \times \boldsymbol{h}_{def}^{disc} & l \in G^{def} \end{cases}, \tag{12}$$

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:

$$\boldsymbol{x}^T = \begin{cases} \boldsymbol{Q}_c \circ \left(\boldsymbol{T}^{pla} - \boldsymbol{T}^{def}\right) \times \boldsymbol{h}_{pla}^{disc} & l \in G^{pla} \\ \boldsymbol{Q}_c \circ \left(\boldsymbol{T}^{def} - \boldsymbol{T}^{pla}\right) \times \boldsymbol{h}_{def}^{disc} & l \in G^{def} \end{cases}. \tag{13}$$

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:

$$\boldsymbol{o}_1 = tanh\left(\boldsymbol{W}_1 \,\text{Concat}\left(\boldsymbol{x}^A, \boldsymbol{x}^T\right)\right), \tag{14a}$$

$$\boldsymbol{o}_2 = tanh\left(\boldsymbol{W}_2 \boldsymbol{o}_1\right), \tag{14b}$$

$$\boldsymbol{y} = sigmoid\left(\boldsymbol{W}_3 \boldsymbol{o}_2\right). \tag{14c}$$

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 [**?**]. 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 $\hat{y}$:

$$\mathcal{L} = -\sum_i \left(y_i \log \hat{y}_i + (1-y_i)\log\left(1-\hat{y}_i\right)\right). \tag{15}$$

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.

The case types are divided into 12 categories and 525 fields. The type in the method proposed in this paper refers to the category and field, that is, a case will be recorded in the category and field it belongs to at the same time. The 10 categories of cases include intellectual property rights, criminal law, ordinary civil disputes, real estate category, contract dispute, objection to jurisdiction, company dispute, enforcement case, administrative dispute, investment and financing business, maritime and merchant shipping, foreign-related category. The website [4] provides tags for the case results, which are divided into 0, 0.5, 0.7, 0.9 and 1. We analyze some judgment documents and tags, and believe that the case results indicate the ratio of winners and losers of the parties. For example, if the party receives most of the compensation, the label is 0.7 or 0.9; The party receives the full compensation, and the label is 1. The party is commuted with a label of 0.7, 0.9, or 1. Therefore, we take the case outcome label as the lawyer benefit label, that is, the benefit that the lawyer is considered to obtain for the client with a label value greater than 0.5.

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
4. https://www.51djl.com/

TABLE 2
Dataset Statistics.

Dataset	LCHR-small	LCHR
# Lawyers	26,430	29,015
# Cases	193,444	772,920
# 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

Fig. 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 [**?**]. 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. If either party of the plaintiff or defendant's team has no participating lawyer, the lawyer with id 0 shall fill in. LawyerPANE adds constant initialization to ensure an initial proficiency parameter of 0.5. For training, the learning rate of Adam optimizer [**?**] 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** [**?**]: 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
Predictions of lawyers' effect on cases. The values in brackets of LawyerPAN indicate relative increase over NeuralCDM-BERT. The values in brackets of variants of LawyerPAN indicate relative decrease compared to LawyerPAN.

Methods	LCHR-small				LCHR			
	Accuracy	RMSE	AUC	F1	Accuracy	RMSE	AUC	F1
AVG	0.706	0.443	0.663	0.810	0.715	0.438	0.685	0.816
LR	0.709	0.459	0.619	0.829	0.694	0.458	0.652	0.819
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	0.875 (+0.13)	0.304 (-0.10)	0.932 (+0.13)	0.909 (+0.08)	**0.889** (+0.13)	**0.286** (-0.11)	**0.946** (+0.13)	**0.920** (+0.10)
LawyerPAN-EI	0.723 (-0.15)	0.451 (+0.15)	0.623 (-0.31)	0.831 (-0.08)	0.727 (-0.16)	0.438 (+0.15)	0.686 (-0.26)	0.831 (-0.09)
LawyerPAN-SC	0.767 (-0.11)	0.393 (+0.09)	0.821 (-0.11)	0.838 (-0.07)	0.782 (-0.11)	0.381 (+0.10)	0.847 (-0.10)	0.844 (-0.08)
LawyerPAN-AM	0.867 (-0.01)	0.313 (+0.01)	0.918 (-0.01)	0.904 (-0.01)	0.880 (-0.01)	0.294 (+0.01)	0.939 (-0.01)	0.912 (-0.01)
LawyerPANE	**0.890** (+0.02)	**0.281** (-0.02)	**0.951** (+0.02)	**0.921** (+0.01)	0.889 (+0.00)	0.287 (+0.00)	0.944 (-0.00)	0.920 (+0.00)

- **BERT** [**?**]: 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** [**?**]: a latent factor model which projects the users and items into a low-dimensional space by mining the sparse consumption matrix.
- **IRT** [**?**], [**?**]: a cognitive diagnosis method, which models students' latent traits and the parameters of exercises like difficulty and discrimination with a logistic-like function.
- **TrueSkill** [**?**]: a rating system among game players. It works well for both teams with different numbers of players.
- **NeuralCDM** [**?**]: 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 removes the case descriptions and only uses the case one-hot embedding to learn the case difficulty.
- **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., $\boldsymbol{h}_{pla}^{diff} = \boldsymbol{h}_{def}^{diff}$, $\boldsymbol{h}_{pla}^{disc} = \boldsymbol{h}_{def}^{disc}$). Therefore, based on formula (10a) and formula (10b), the context module on formula (13) can be reformulated as:

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

TABLE 4
Experimental results on uniformization function.

Methods	Accuracy	RMSE	AUC	F1
# MeanPooling	0.875	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 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.

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 only adopts the anticipatory module to assess the proficiency of lawyers. It performs poorer than LawyerPAN, which not only demonstrates the importance of the case factors while using text information and the maximum difference of the proficiencies of all lawyers, but also shows the adversarial module takes an active role in proficiency assessment for lawyers. 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

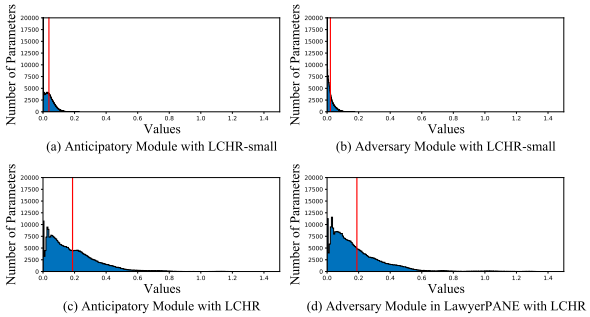


Fig. 5. Visualization of interaction parameter.

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 (19) and (3) respectively. The more precise performance of LawyerPAN-AM proves the effectiveness of joint probability formula (3).

5.4.3 *Visualization of interaction parameter.*

As mentioned before, LawyerPAN-AM has poorer performance than LawyerPAN, which indicates that the modeling of adversarial module is more effective for proficiency assessment. To further show this conclusion, we analyze $\boldsymbol{W}_1$ in formula (14) of the models which both have anticipatory and adversarial modules using LCHR-small and LCHR datasets. Specifically, $\boldsymbol{W}_1$ contains values of the weight of representations of each module.

We perform the distribution of parameters in Figure 5, where the red line denotes the average value. From the observations, we can find that the most value of parameters in the anticipatory module is higher than the adversarial module, either in LCHR-small and LCHR datasets. Owing to the positively correlation which is introduced in section 4.3, we can get a conclusion that the higher the model weight, the greater the impact on the final effectiveness. Therefore, we regard the anticipatory module has more effectiveness in proficiency assessment.

5.4.4 *Accuarcy of each category.*

We counted the prediction accuracy of various types of cases of the processed data, and the results are shown in Table 5. The results of intellectual property, civil disputes and contract disputes were preferred, while the prediction accuracy of administrative disputes was low, which we guessed was caused by the different number distribution of various types of cases in the data set.

5.4.5 *Impact of uniformization function.*

Afterward, due to the impact of the adversarial module on proficiency assessment, we turn to discuss this adversarial module. Specifically, in the practical litigation context, debates in court are usually conducted in teams. Intuitively, the comprehensive proficiency of the lawyer teams depends on the lawyer with the highest proficiency to a large extent and also the influence of other lawyers. Therefore, we make a series of experiments about the variants of formula (11)

TABLE 5
Accuarcy of each category.

Category	Accuarcy
civil dispute	0.903
contract dispute	0.819
company dispute	0.906
intellectual property right	0.960
administrative dispute	0.790
maritime and merchant shipping	0.813

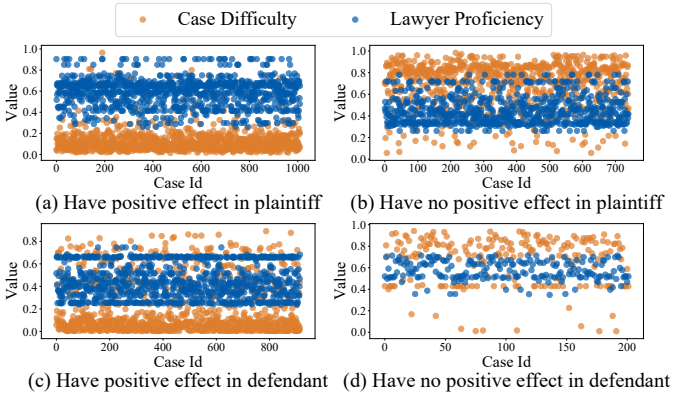


Fig. 6. Visualization of assessment parameters.

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 max-pooling approach has a better result than min-pooling and mean-pooling, which demonstrates the importance of the lawyer with higher proficiency. Meanwhile, different from the above methods, the formula (11) we designed considers not only the impact of the most proficient lawyer, but also the influence of other lawyers. Therefore, at the same time of getting the best result, it is also consistent with our intuitive understanding.

5.4.6 *Impact of IRT parameter.*

We tried 1-parameter, 2-parameter, and 3-parameter IRT, respectively, expressed as follows:

- 1-parameter IRT: a constant of 0.5 is used for case difficulty.

$$P(\theta) = \frac{1}{1 + e^{-D\alpha(\theta - 0.5)}}, \tag{17}$$

- 2-Parameter IRT: The underlying IRT approach adopted by LawyerPANE in this paper.

$$P(\theta) = \frac{1}{1 + e^{-D\alpha(\theta - \beta)}}, \tag{18}$$

- 3-parameter IRT: The guessing index c is introduced.

$$P(\theta) = c + (1 - c)\frac{1}{1 + e^{-D\alpha(\theta - \beta)}}, \tag{19}$$

The experimental results are shown in the Table 6, and the experimental results show the effectiveness of the lawyer ability evaluation method we adopted by fusing Q matrix and neural network.

TABLE 6
Experimental results on IRT parameter.

Methods	Accuracy	RMSE	AUC	F1
# 1-parameter	0.753	0.450	0.784	0.821
# 2-parameter	0.890	0.281	0.951	0.921
# 3-parameter	0.695	0.461	0.500	0.820

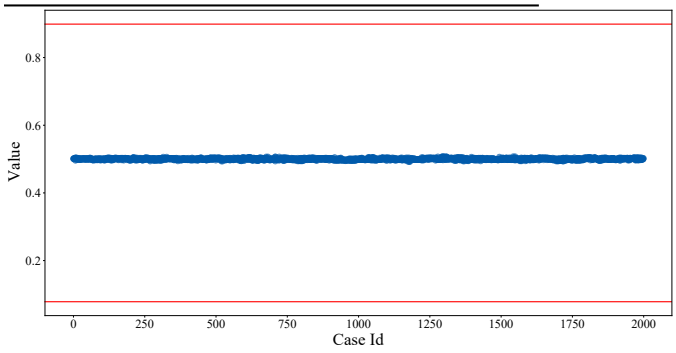


Fig. 7. Visualization of assessment parameters.

5.4.7 *Visualization of assessment parameters.*

To compare with the lawyer proficiency, our LawyerPAN model evaluate the difficulty of the cases and Figure 6 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.

5.4.8 *Visualization of proficiency scope in LawyerPANE.*

We believe that the scope of the lawyer's competence depends on the difficulty of the case, that is, greater than the lowest difficulty of the failed case and less than the highest difficulty of the successful case. Because below the minimum difficulty of losing cases that all lawyers can win, we can assume that the lawyer has little ability to win any lawsuit, even lower than non-law practitioners. And above the highest difficulty of a successful case where none of the lawyers who take this case can win the lawsuit, we can say that the lawyer has basically achieved the highest level in the field. The red line of the graph shows the highest and lowest difficulty of the successful case, the blue dot shows shows the proficiency of this lawyer in Figure 7.

5.5 Exploring the Representation of Lawyers and Cases

We will discuss the representation of lawyers' proficiency, the cooperative and competitive relationship between lawyers, and the representation of the case in the plaintiff and defendant. The results are shown in Table 7.

5.5.1 *General proficiency and Professional proficiency of Lawyers*

Generally speaking, the lawyer's proficiency is divided into **G**eneral proficiency and **P**rofessional proficiency. General proficiency refers to the level of proficiency that most of lawyers can achieve, and professional proficiency refers to the level of proficiency that a lawyer exceeds other lawyers in an area of expertise. We use the sum of the two to represent a lawyer's proficiency. Therefore, we define the lawyer proficiency as:

representation $\boldsymbol{x}^l$. That is,

$$\boldsymbol{h}^l = \text{sigmoid}\left(\left(W\boldsymbol{x}^l + q\right) \times \boldsymbol{E}^l\right), \tag{20}$$

where $x^l \in \{0,1\}^{1\times N}$ is one-hot representation of the lawyer, matrix q represents general proficiency of all lawyers, trainable matrix $\boldsymbol{E}^l \in \mathbb{R}^{N\times K}$.

Table 7 shows the experiments on the representation of lawyers. The results demonstrate the validity of dividing lawyer's proficiency into general and professional proficiency.

5.5.2 *cooperative and competitive relationship between lawyers*

Cooperation and **C**ompetition exist in the process of attorney-represented litigation. For example, lawyer A is aware of lawyer B's professional weakness, and lawyer A may argue with lawyer B about this during the trial. Or lawyer A and lawyer B often work together, and their cooperation can greatly improve the probability of the agent obtaining the rights. Therefore, the modeling of cooperation and competition is very important in the litigation process. In this report, we define the anticipatory module of the assessment layer as:

$$\boldsymbol{x} = \boldsymbol{Q}_c \circ \left(\boldsymbol{h}^l + \boldsymbol{h}^r - \boldsymbol{h}^{\text{diff}} \times \left(1 + \boldsymbol{h}^{fac}\right)\right) \times \boldsymbol{h}^{disc}, \tag{21}$$

where $\boldsymbol{h}^r$ denotes the cooperation and competition of lawyers.

Then, we explore the modeling of collaboration and competition using the following three approaches:

- LPAN-C&C-A:

$$\boldsymbol{h}^r = \boldsymbol{h}^s \circ \left(\boldsymbol{x}^l \times \mathbf{F}\right), \tag{22}$$

  where matrix F represents the every correlation of cooperation and competition between two lawyers.
- LPAN-C&C-B:

$$\boldsymbol{h}^r = \boldsymbol{x}^l \times \mathbf{F} \times \boldsymbol{E}^l, \tag{23}$$

- LPAN-C&C-C: Considering that the matrix F is a symmetric matrix, we optimize the formula as the follow:

$$\boldsymbol{h}^r = \boldsymbol{x}^l \times \mathbf{F}^T\mathbf{F} \times \boldsymbol{E}^l, \tag{24}$$

Table 7 shows the experiments on cooperation and competition of lawyers. The results show that the factors of

cooperation and competition are not considered as well as LPAN-base model, which may be because the representation is not sufficient for interaction. We will study the modeling of cooperation and competition in the future.

5.5.3 the difficulty of Plaintiff and Defendant cases

Although there is a distinction between the difficulty of **P**laintiff and **D**efendant cases in our work, we have tried some other modeling methods. In this report, we define the anticipatory module of the assessment layer as the following two formulas:

- LPAN-P&D-A:

$$\begin{aligned} \boldsymbol{x} &= \boldsymbol{Q}_c \circ \left(\boldsymbol{h}^l - \boldsymbol{h}^{\text{diff}} \times \left(1 + \boldsymbol{h}^{fac}\right)\right) \times \boldsymbol{h}^{disc}, \\ \boldsymbol{h}^{\text{fac}} &= \text{sigmoid}\left(\boldsymbol{x}^f \times \mathbf{E}\right), \end{aligned} \tag{25}$$

  where $\boldsymbol{h}^{fac}$ denotes the faction of lawyers, i.e., $\boldsymbol{h}_i^{fac} = 0$ if the lawyer agent for the plaintiff and $\boldsymbol{h}_i^{fac} = 1$ otherwise.
- LPAN-P&D-B: We consider that there might be a relationship between co-operative competition and legal competence, so we tried to represent both:

$$\boldsymbol{x} = \boldsymbol{Q}_c \circ \left(\boldsymbol{h}^s + \boldsymbol{h}^r - \boldsymbol{h}^{diff} \times \left(1 + \boldsymbol{h}^{fac}\right)\right) \times \boldsymbol{h}^{\text{disc}}, \tag{26}$$

Table 7 shows the experiments on the representation of cases. According to the experimental results and the research [2], modeling plaintiff and defendant cases separately can significantly improve the performance of the method, and we will refine the case modeling in the future

TABLE 7
The experiments on the representation of lawyers and cases.

Methods	LCHR-small		
	Accuracy	RMSE	AUC
LPAN-G&P	0.809	0.377	0.778
LPAN-C&C-A	0.784	0.399	0.702
LPAN-C&C-B	0.792	0.388	0.701
LPAN-C&C-C	0.770	0.403	0.659
LPAN-P&D-A	**0.821**	**0.361**	**0.823**
LPAN-P&D-B	0.803	0.380	0.769

5.6 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.

5.6.1 *Individual perspective.*

We give an example of the assessment reports containing the individual lawyers' proficiency in each legal field in Figure 8. Here, we demonstrate the rationality and effectiveness

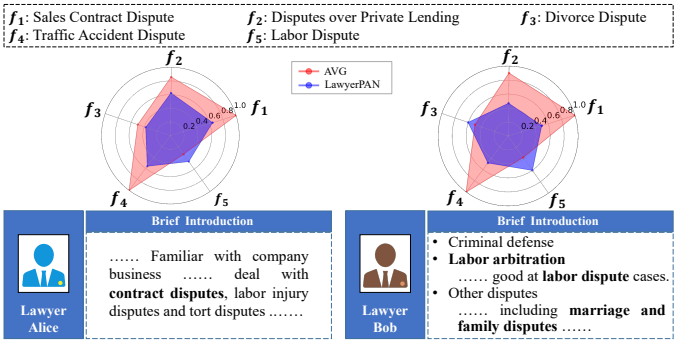


Fig. 8. An assessment example of 2 lawyers. The radar figure shows proficiency assessment on 5 common fields with similar average effectiveness. Bottom descriptions are excerpted from the resumes on their home page of websites.

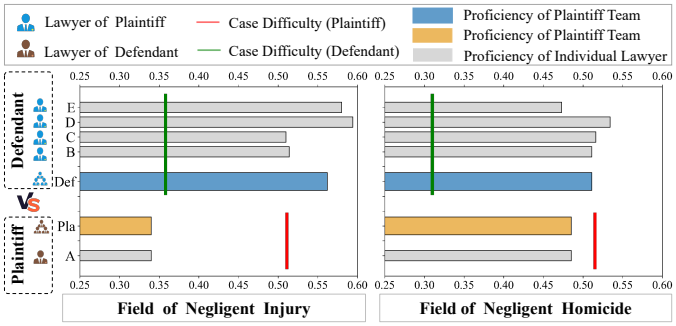


Fig. 9. A case study of lawyer assessment in the adversarial module of LawyerPAN.

of LawyerPAN from two aspects. On the one hand, in the field f_1 (i.e., *Sales Contract Dispute*), since both lawyers have 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 Bob is more proficient than lawyer Alice 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 Bob. In this sense, the LawyerPAN is in line with the reality.

5.6.2 *Team perspective.*

Furthermore, to intuitively verify this interpretability, we design a visualization in Figure 9 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 Discussion

6.1 Ethic Discussion

From the above experiments, we can observe that LawyerPAN provides both accurate and interpretable results for proficiency assessment of lawyers. There still some challenging problems for future studies. For approaches, we will focus on changeable proficiency and a more precise separation of proficiency level. For the data, data sparsity and more complicated situations are challenging problems. Then, we futher discuss the ethical issues concerning our work. Generally, the lawyer aims to achieve the best possible outcome for his client (i.e., litigants). However, lawyers may focus on his own benefits because of the needs of their job, thus make defense actions that run counter to legal fairness, even if they do not intend to. Therefore, we emphasize most strongly that the purpose of our modeling is to assess the professional ability of lawyers to protect the rights of clients, not only for the professional development of lawyers themselves, but also for avoiding unjust, false, and wrong judgment results to the greatest extent, and maintaining the fairness and justice of the law.

6.2 Limitation

This paper explores the ability evaluation of litigators in legal litigation scenarios, while non-litigators are more likely to write legal related documents, which is difficult to evaluate only through legal litigation results.

This paper proposes a proficiency assessment method for litigation lawyers, which aims to relatively assess the proficiency level of lawyers in various fields. In the future, it is planned to use the proficiency assessment results of lawyers for personalized services, such as case recommendation and development planning. However, for the evaluation of a lawyer in the actual legal scene, more actual situations and factors need to be combined, and this method can only provide reference value for the evaluation of a lawyer's proficiency.

From the method level, our proposed method still needs to be improved on the cold start problem, specifically, when a new lawyer or a new case is added, the proposed method needs to retrain the model.

6.3 Extensions

Inspired by the cognitive diagnosis in the field of education [**?**], this paper improves LawyerPAN [2]. The representation part of case difficulty can still be more fine-grained, including considering more detailed exploration of the case text structure [**?**], the charge relationship [**?**], or the dialogue information [3] shown in Figure 10.

6.4 Future work

The next steps include: 1) further strengthen data cooperation, obtain more dimensional data of lawyers, including more fine-grained field data of lawyers, judgment documents of accepted cases, performance completion, etc., and

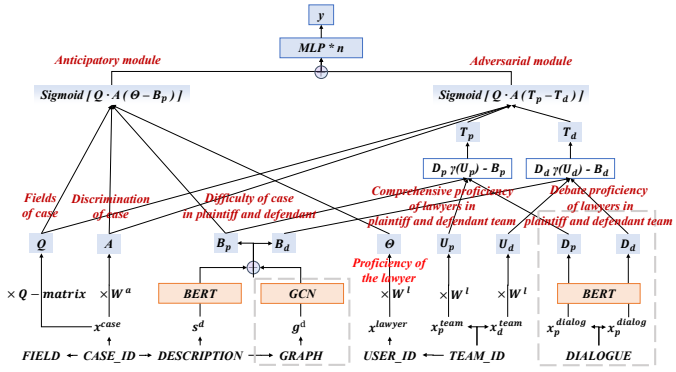


Fig. 10. The extension of the proposed LawyerPANE model. Overall, from the bottom up, it can be divided into 3 layers: 1) Input and Embedding Layer, which consists of assessment lawyers, cases and team factors; 2) Assessment Layer, which contains two modules to learn multiple complex relations between lawyers and cases; 3) Interaction and Output Layer, which is designed for feature learning and final prediction.

improve the accuracy of evaluation of lawyers 'professional ability and specialties. 2) The algorithm framework is extended to the professional proficiency evaluation of other legal service practitioners, including forensic experts, notaries, mediators, etc.

7 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). The calculations in this work are supported by the Supercomputing Center of USTC.

References

[1] L. Gong, X. Feng, D. Ye, H. Li, R. Wu, J. Tao, C. Fan, and P. Cui, "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*, 2020, pp. 2300–2310.

[2] Y. An, Q. Liu, H. Wu, K. Zhang, L. Yue, M. Cheng, H. Zhao, and E. Chen, "Lawyerpan: A proficiency assessment network for trial lawyers," in *Proceedings of the 27th ACM SIGKDD conference on knowledge discovery & data mining*, 2021, pp. 5–13.

[3] Y. An, L. Yue, W. Gao, K. Zhang, and Q. Liu, "Post-hoc facts augmented legal judgment prediction," in *International Conference on Database Systems for Advanced Applications*. Springer, 2024, pp. 476–485.
[4] J. Becker, J. P. Wahle, B. Gipp, and T. Ruas, "Text generation: A systematic literature review of tasks, evaluation, and challenges," *Journal of Artificial Intelligence Research*, vol. 86, 2026.
[5] L. Yue, Q. Liu, H. Wu, Y. An, L. Wang, S. Yuan, and D. Wu, "Circumstances enhanced criminal court view generation," in *Proceedings of the 44th international ACM SIGIR conference on research and development in information retrieval*, 2021, pp. 1855–1859.
[6] H. Ye, X. Jiang, Z. Luo, and W. Chao, "Interpretable charge predictions for criminal cases: Learning to generate court views from fact descriptions," in *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, 2018, pp. 1854–1864.
[7] Y. Wu, K. Kuang, Y. Zhang, X. Liu, C. Sun, J. Xiao, Y. Zhuang, L. Si, and F. Wu, "De-biased court's view generation with causality," in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2020, pp. 763–780.