Risk Analysis for Intellectual Property Litigation

Mihai Surdeanu, Ramesh Nallapati, George Gregory, Joshua Walker
Lex Machina
{mihai,nmramesh,george.gregory,jwalker}@lexmachina.com

Christopher D. Manning
Stanford University
manning@stanford.edu

ABSTRACT

We introduce the problem of risk analysis for Intellectual Property (IP) lawsuits. More specifically, we focus on estimating the risk for participating parties using solely prior factors, *i.e.*, historical and concurrent behavior of the entities involved in the case. This work represents a first step towards building a comprehensive legal risk assessment system for parties involved in litigation. This technology will allow parties to optimize their case parameters to minimize their own risk, or to settle disputes out of court and thereby ease the burden on the judicial system. In addition, it will also help U.S. courts detect and fix any inherent biases in the system.

We model risk estimation as a relational classification problem using conditional random fields [6] to jointly estimate the risks of concurrent cases. We evaluate our model on data collected by the Stanford Intellectual Property Litigation Clearinghouse, which consists of over 4,200 IP lawsuits filed across 88 U.S. federal districts and ranging over 8 years, probably the largest legal data set reported in data mining research. Despite being agnostic to the merits of the case, our best model achieves a classification accuracy of 64%, 22% (relative) higher than the majority-class baseline.

1. INTRODUCTION

Intellectual Property (IP) law handles legal property rights over creations of the mind, such as industrial, literary and other artistic works. Enforcing IP rights is crucial for industries where the cost of replicating a product is significantly smaller than the cost of creating that good (*e.g.*, pharmaceutical). Consequently, IP lawsuits may have drastic consequences (*e.g.*, in a patent infringement case in the pharmaceutical industry, the damages paid by the defendant run into millions of dollars if infringement is proved). Hence, it is a matter of critical importance for parties involved in IP litigation to continually assess their respective risks during the entire progression of the case, starting from its filing time, or even prior to filing. Understanding these risks will help the parties decide whether to continue to fight the case in the hope of a favorable jury verdict, or to settle out of court, thus avoiding large litigation expenses and maintaining the possibility of negotiating the outcome. As a desirable side-effect, if most parties settle cases out of court using an accurate risk analysis system, it would also help ease the burden on the judicial system and make it more efficient.

The main goal of our research is to build a risk analysis model for IP litigation. Risk analysis is a complex problem that depends on a number of factors. Broadly, we classify most factors into one of two categories: (a) *merits of the case* – factors in this category include the strength of the patents asserted, similarity of the defendant's manufacturing technology to the patent's technology, *etc*; and (b) *prior factors* – factors that do not model directly the merits of the case but instead focus on past information that may influence the outcome of the current case. For example, this latter category includes the past win rates of parties, attorneys and law firms involved in the case, potential biases of judges estimated from past cases, *etc*. While both these classes of factors are important for risk estimation, in this paper we focus on the latter category. More specifically, this work raises the following questions:

1. Are prior factors relevant in determining the risks involved in a case?
2. What are the important prior factors and what is their relative significance in estimating the overall risk of a case?

Focusing on prior factors is important for several reasons. Firstly, prior factors model ingredients that may influence the case outcome independently of the case merits, *e.g.*, a good attorney may increase the chances of winning. Understanding what these factors are provides practical feedback for risk minimization: at the time of litigation, parties cannot change the merits of the case but can adjust prior factors to optimize chances of winning. Secondly, prior factors indirectly model the merits of the case. For example, the fact that a plaintiff has a high success rate in previous litigation is a likely indication that it owns strong patents. Hence, prior factors may prove useful also for modeling the merits of cases. And lastly, prior factors are easier to extract than the merits of the case which may require more sophisticated modeling.

In this work we answer the above questions empirically using a *prior risk analysis model*, *i.e.*, a model that uses only prior factors for estimation, as follows. To answer the first question on the influence of prior factors for the estimation of litigation risk, we model prior risk as a binary classification task, *i.e.*, patent-owner (plaintiff) wins or defendant wins. The risk for any party can be estimated in terms of the probability that the opposing party wins. In the prediction model we use only prior factors as features, *i.e.*, prior performance of the entities involved in the case, which can be extracted at the time the corresponding case was filed. We train and evaluate our model using data provided by the Stanford Intellectual Property Litigation Clearinghouse (IPLC) project[1], which covers IP litigation from the United States from the past eight years. Our results indicate that such models have a prediction accuracy of approximately 64%, which is significantly higher than the majority-class baseline. To our knowledge, this work is the first to show that litigation risk can be estimated using solely prior information.

[1] `http://www.law.stanford.edu/program/centers/iplc/`

ICAIL '11, June 6-10, 2011, Pittsburgh, PA

	Party	Attorney	Law firms
Frequency	36.1%	65.4%	90.5%
Same outcome	80.9%	69.5%	61.4%

Table 1: Analysis of concurrent lawsuits. The first row indicates the percentage of lawsuits in the training corpus that share at least one attribute of the given type (*i.e.*, party, attorney, or law firm) in the same role (plaintiff or defendant) with at least another concurrent lawsuit. The second row indicates the percentage of connected cases that have the same outcome, from the corresponding subsets listed in the first row.

To answer the second question on the relative importance of each feature, we performed an extensive feature analysis of our model. This analysis indicates that the prior performance of the attorneys and law firms involved in the case is a crucial feature for prediction. Furthermore, we show that there is significant correlation between concurrent cases that share parties. We exploit this correlation using statistical relational learning to further improve the predictive power of our model.

2. APPROACH

We model the prior risk analysis problem as a supervised discriminative binary classification task whose goal is to predict the outcome of new litigation given relevant prior factors. In this paper we focus on patent-infringement cases that had a publicly known polarized outcome, *i.e.*, cases that were not settled. We model each lawsuit as a separate datum that is assigned one of two possible labels: *patent owner wins* or *accused defendant wins*. For simplicity, in the remainder of the paper we will refer to the patent owner as the plaintiff.[2] We introduce next the models we built for this task. We conclude this section with the description of the features that implement the prior factors in our models.

2.1 Models

Each individual lawsuit is modeled as a distinct example x_i in our dataset containing n examples. An example x is represented as a vector of features $\mathbf{f}(x) = (f_1(x), \cdots, f_m(x))$ where m is the number of distinct features. In the simplest representation, features are extracted using solely historical information, *i.e.*, information extracted from cases that terminated before the case to be modeled was filed. Each component $f_j(x)$ could be a binary value, an integer, or a real valued fraction.

We use the $L2$ regularized logistic regression with the following objective function as the classifier:

$$P(\mathbf{y}|\mathbf{x}, \mathbf{w}, \sigma^2) = (\prod_{i=1}^{n} \frac{\exp(y_i \mathbf{w}^T \mathbf{f}(x_i))}{1 + \exp(\mathbf{w}^T \mathbf{f}(x_i))}) \frac{\exp(-\frac{\mathbf{w}^T \mathbf{w}}{2\sigma^2})}{\sqrt{2\pi\sigma^2}} \quad (1)$$

where $y \in \{1, 0\}$ represents the outcome label of the case,[3] $\mathbf{w}$ is a vector of feature specific weights, and σ^2 is a regularization parameter that can be tuned.

A limitation of the above model is that it cannot represent dependencies between concurrent[4] and correlated cases. Intuitively, concurrent cases that share information *should* have a correlated outcome. For example, two concurrent cases with the same company as plaintiff are likely to be on the same topic, *i.e.*, infringement of related patents, use the same evidence and, hence, have the same outcome. This assumption is verified empirically in Table 1, which analyzes the correlation between concurrent cases in our training corpus (the corpus is detailed in the next section). The first row in the table shows the percentage of cases in the training corpus that share one or more entities in the given role with at least another concurrent case. Note that these values are generally high because there is significant chronological overlap between cases: because the average time to termination of an IP case is larger than one year, virtually every case in our corpus overlaps with other cases. The data indicates that sharing is most prevalent for law firms and least common for parties, *e.g.*, 90% of cases have one or more concurrent case with the same law firm in the same role but only 36% of cases share a party in the same role with another concurrent lawsuit. This is to be expected because law firms participate in significantly more cases than individual attorneys or parties. However, the second row in the table shows that the correlation between outcomes is strongest for cases that share parties: in the set of cases that share at least one party with one or more concurrent cases, over 80% have the same outcome as the corresponding concurrent cases.

The above analysis suggests that there is merit in modeling the correlation between concurrent cases, especially for cases that share a party in an identical role. However, this correlation does not always hold: 20% of the concurrent cases that share parties have different outcomes. So, imposing a hard constraint at runtime that any concurrent cases must have the same outcome would be a bad idea. Instead, we prefer to learn soft constraints that estimate the strength of correlation between concurrent cases using statistical relational learning. In this paper, we use conditional random fields (CRF), implemented over the network of concurrent lawsuits. To the best of our knowledge, this work is the first to show that statistical relational learning is applicable to the problem of litigation risk analysis.

Formally, we first define a graph $G = (V, E)$ as follows. Each vertex $v_i \in V$ corresponds to a case x_i in the data. We also define an edge $e_{ij} \in E$ between vertices v_i and v_j if and only if x_i and x_j are concurrent cases that share at least one party in the same role (*i.e.*, the shared entity occurs either as plaintiff or as defendant in both cases). We now define a CRF over the graph G that jointly models the outcomes $\mathbf{y} = (y_1, \cdots, y_n)$ of all cases $\mathbf{x} = (x_1, \cdots, x_n)$ as follows:

$$P(\mathbf{y}|\mathbf{x}) = \frac{1}{Z} \exp(\sum_{i=1}^{n} y_i \mathbf{w}^T \mathbf{f}(x_i) + \sum_{(i,j) \in E} \mathbf{v}^T \mathbf{g}(y_i, y_j)) \quad (2)$$

where $\mathbf{g} = (g_0, g_1)$ are our new network features, defined as $g_0(y_i, y_j) = 1$ if $y_i = y_j$ and 0 otherwise, and $g_1(y_i, y_j) = 1 - g_0(y_i, y_j)$. In other words, g_0 is active if the two cases x_i and x_j have the same outcome and g_1 is active if they have opposite outcomes. The weight vector $\mathbf{v}$ corresponding to these features captures the network correlation strength.

Learning:
Although this model is attractive for the problem of joint modeling of concurrent cases, exact learning is intractable for an arbitrary graph such as the one we defined in our problem. Hence, in this paper we use a variant of pseudo-likelihood for training [3]. Pseudo-likelihood is known to be a consistent estimator of true likelihood and is known to work well in cases where local features are strong. In this method, the joint likelihood of all the variables in a model is approximated by the product of the probability of each variable, conditioned on all other variables as shown below:

[2] This is not always the case: in declaratory judgment cases the accused party initiates a non-infringement lawsuit, hence the patent owner is legally the defendant. This happens when a party is threatened with an infringement lawsuit by the patent owner but that lawsuit is not yet filed. In this paper, we normalize these notations to their semantic interpretation, *i.e.*, the patent owner is always the alleged prejudiced party or the plaintiff, regardless of who filed the first complaint.

[3] '1' represents plaintiff wins and '0' defendant wins.

[4] Two cases are concurrent if neither terminates before the other starts.

$$P(\mathbf{y}|\mathbf{x}) \approx \prod_i P(y_i|\mathbf{y}_{-i}, x_i) = \prod_i P(y_i|\mathbf{y}_{N(i)}, x_i) \quad (3)$$

where the subscript $_{-i}$ refers to all variables not including y_i, and $N(i)$ refers to the neighbors of y_i. The second step above follows from the fact that a variable is conditionally independent of all other variables given its neighbors, in an undirected graphical model. In the case of an exponential model such as the CRF, each term in the product above would be equal to the following logistic regression function:

$$P(y_i|x_i, \mathbf{y}_{N(i)}) = \frac{1}{Z'} \exp(y_i \mathbf{w}^T \mathbf{f}(x_i) + \sum_{j \in N(i)} \mathbf{v}^T \mathbf{g}(y_i, y_j)) \quad (4)$$

This is much easier to learn than the joint model because it requires no global information propagation.

Inference:
Since exact inference is computationally expensive as well, we use Gibbs sampling [1] to perform approximate inference. Similar to pseudo-likelihood, Gibbs sampling deals with the same local probability $P(y_i|x_i, \mathbf{y}_{N(i)})$ shown in the last equation. In this approach, we sample each variable y_i in turn from the local probability, where $\mathbf{y}_{N(i)}$ correspond to the latest outcome assignments of its neighbors. This iterative process, when run long enough, is guaranteed to converge to the true posterior.

Also, since we need best variable assignments rather than true posterior, we use simulated annealing with Gibbs sampling, using a linear cooling schedule, as used in [5]. According to this approach, we exponentiate the sampling distribution in the last equation by a value $1/C$. We initialize $C = 1$ at the start of Gibbs iterations, and we decrease it linearly with each iteration until $C \rightarrow 0$. At small values of C, the sampling distribution becomes peaked at the maximizing value, thus returning the maximal assignments.

2.2 Features

We model the past behavior of all litigation entities involved in a given case. These entities are: (a) the *parties* involved in the case, *i.e.*, plaintiffs and defendants; (b) the *attorneys* on each side of the lawsuit, *i.e.*, plaintiffs' and defendants' attorneys; (c) the plaintiffs' and defendants' *law firms*; (d) the *judges* assigned to a case (usually, there is a single judge per case, but a lawsuit may eventually have multiple judges if it is transferred to a different district); and (e) the *districts* where the case is filed.

It is obvious that parties, attorneys, or law firms are important parameters of a lawsuit. Judges are also important because they decide what information is presented to the jury. Hence, their own bias may indirectly influence jury decisions. The district is important because the jury is selected from the local population, which may have certain cultural biases (either pro or against IP). We model the past behavior of the entities listed above using four different types of features:

(a) *Unique identifier* (`id`) – for each participating entity in a given case we generate a Boolean feature for its unique identifier concatenated with its role in the case. For example, the feature `id:plaintiff-attorney:101` indicates that the entity with id `101` served as a plaintiff's attorney in the current case. This allows the discriminative model to learn the correlation between outcome labels and entities in a given role. For example, if the previous feature appears mostly in lawsuits where the plaintiff side wins, this indicates that the corresponding attorney is usually successful when defending the plaintiff side.

(b) *Past win rates in any role* (`wr`) – we model the win rates of lawsuit entities, *regardless of their role* in past lawsuits. We compute this feature explicitly for each entity as the percentage of past cases, *i.e.*, cases that terminated before the current case and included the given entity in any role, that were won by the side of the corresponding entity. For example, if an entity won a past case as plaintiff, another as defendant, and lost one as plaintiff, its win rate is 66%. Since there are multiple entities that have the same role in a given case, we average all win rates for the entities in a given role and use this average as the actual feature for that role. These two feature sets (`id` and `wr`) are intended to capture both the dependency between previous successes and the current outcome (*e.g.*, a better law firm should increase changes of winning) but also to model, albeit indirectly, the merits of the case (*e.g.*, a party with a high success rate in previous litigation is likely to own strong IP, which should be reflected in the outcome of the current case as well). These features are not computed for judges and districts.

(c) *Judge and district bias* (`bias`) - for judges and districts we compute a variant of the `wr` feature that estimates the bias of that judge or district towards one side of the litigation. This feature is computed as the ratio of cases won by the plaintiff from the set of past cases assigned to the corresponding judge or district. Same as above, we use as the actual feature value the average over all judges or districts assigned to the case.

(d) *Counts of participation in past cases in any role* (`count`) – this feature counts the presence of the corresponding entity in past lawsuits, regardless of its role and the outcome of that case. For example, the value of this feature for the entity in the above `wr` example is 3, because this entity participated in three past cases. This feature serves as an estimate of experience. In other words, the amount of litigation experience that an entity has is likely to be correlated with the number of cases in which it has participated. Same as the `wr` feature, this feature is averaged over all entities of same type and is not computed for judges and districts.

We computed the values of all non-boolean feature types, *e.g.*, `wr`, `bias`, and `count`, using only historical information, *i.e.*, information gathered from lawsuits that terminated before the current case was filed.[5] Barring the exceptions noted above, we generate all combinations of feature types and entity types. This yields a total of 22 distinct feature groups.

3. DATA

The data used in this paper was provided by the Stanford IPLC project. The corpus consists of all IP lawsuits between beginning of 2001 and end of 2008. The cases in the corpus were previously annotated with their outcomes. The annotation process followed a pipeline model: first, two IP experts generated the initial annotations; second, an IP attorney reviewed all outcomes and decided the final annotation.

The meta data available for each lawsuit stores filing and termination times, and the names of all the entities involved (parties, attorneys, law firms, districts and judges). However, all names in the corpus are just textual mentions that maintain the spelling used by the person who filed or registered the transfer of the corresponding case. To transform this data into usable information we implemented an entity resolution (ER) component that consolidates all entity mentions into a set of clusters, where each cluster contains all mentions that point to the same real-world legal entity, e.g., "Quinn Emanuel, LLP" and "Quinn Emanuel Urquhart" are different versions of the same law firm name.

The entire corpus contains 20,980 annotated cases. From this set we discarded 16,666 cases that settled, *i.e.*, the outcome is not

[5]This means that, for some test cases, these feature values include information from other lawsuits in the testing partition. This is closer to our envisioned application, where we estimate the risk of a new case using all the cases already terminated.

Cases	Parties	Attorneys	Law firms	Judges	Districts
4,263	12,270	15,706	5,261	1276	88

Table 2: Summary statistics for the litigation corpus. Only terminated cases with a polarized outcome are considered.

	Plaintiff	Defendant
Parties	4,199	8,852
Attorneys	9,008	10,629
Law firms	2,946	3,625

Table 3: Number of unique entities per type $\times$ role for the 4,263 polarized cases in the corpus.

(publicly) in favor of one of the sides, and 51 cases that had outcomes in favor of both sides. The latter situation happens because outcomes are assigned to lawsuit claims, rather than the entire case. Because this work treats cases as indivisible units, we ignore these situations. The remaining 4,263 cases form the corpus used here.

Table 2 shows the overall statistics for the 4,263 cases. The table shows that, even though the number of cases is relatively small, the corpus contains a significant number of entities, *e.g.*, there are more than 12,000 distinct parties involved. Table 3 shows the relevant statistics per role (plaintiff or defendant). Note that the numbers per entity type (party, attorney, or law firm) do not add to the values reported in Table 2 because some entities appear in different roles in different cases, hence they are counted twice in this table. Table 3 indicates that there are significantly more entities on the defendant side than on the plaintiff side.

Because the core of our prediction model is based on historical information we train and test our approach on cases where at least one out of the six polarized entities (*i.e.*, party, attorney, or law firm on the plaintiff or defendant side) has participated in three or more previous lawsuits. We choose a minimal history of three previous cases because this allows the model to learn unambiguous information about past success rates. This reduces our corpus to 3,243 cases, or 76% of the original corpus. To construct the training and testing partitions we sort all cases in chronological order of their termination date and reserve the first 70% (2,270 cases) for training and the remaining 30% (973 cases) for testing.

4. EXPERIMENTS

4.1 Overall Results

Table 4 shows overall results for three models: a baseline that predicts the majority class (plaintiff wins in our corpus), the logistic regression (LR) model that uses only historical features, *i.e.*, Equation 1, and the CRF model that models concurrent cases jointly, *i.e.*, Equation 2. For both models we report average results over 100 different samples of the test corpus, generated with bootstrap resampling. We compute statistical significance using two-tailed paired t-test at 99% confidence interval on these 100 samples. The table shows that both our models outperform significantly the proposed baseline. Our best model (CRF), improves the baseline with over 22 relative percentage points. These results provide empirical proof that prior factors influence the outcome of IP litigation.

Table 4 also shows that capturing the correlation between concurrent cases is beneficial. The CRF model that incorporates these correlations has statistically-significant improvements over the LR model that exploits only historical information. We hypothesize that this is caused by the repetition of similar evidence in correlated litigation. For example, a plaintiff party that participates in two concurrent cases on the infringement of the same patents, will provide the same evidence in both lawsuits. Hence, it is very likely that these lawsuits will have the same outcome. Repeated information likely appears in other similar (but not identical) cases as well, *e.g.*, different patents owned by the same party. Such correlations are modeled implicitly in our approach, through the sharing of litigating entities between cases.

	Baseline	LR	CRF
Test	52.44	63.36^*	$\mathbf{64.03}^*$

Table 4: Classification accuracy of the two models and the majority-class baseline. The scores suffixed with * are significantly better (2-tailed paired t-test, $p = 0.01$) than the previous value in the same row.

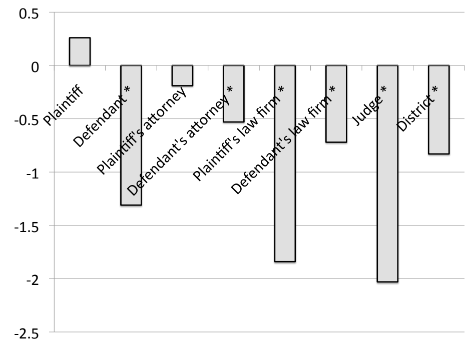

Figure 1: Ablation experiment for the CRF model, with feature groups defined around participating entity types. The feature groups suffixed with * have a statistically-significant impact (2-tailed paired t-test, $p = 0.01$).

4.2 Ablation Analysis

To understand the important factors in our prediction model, we performed several *post-hoc* ablation experiments using our best CRF model. Figure 1 shows the results of an ablation experiment where feature groups are defined around entity types, *e.g.*, we group all the features about plaintiff's attorneys into a single set. The experiment measures the accuracy of the CRF model as each one of these feature groups is removed in turn from the complete feature space. The figure shows the difference in accuracy scores between the system without the corresponding feature group and the best system that uses all features. Hence, negative values in the figure indicate that the corresponding feature group is useful. We draw several conclusions from this experiment:

(a) Legal entities, *i.e.*, attorneys and law firms, are crucial. For example, removing the plaintiff party does not yield a statistically-significant change in performance, but eliminating the plaintiff's law firm causes a drop in accuracy of almost 2 percentage points. On the defendant side, the combined contribution of the defendant's law firm and attorney equals that of the defendant party. Three out of four of the legal entities (law firms on both sides and the defendant's attorney) have a significant impact, with drops in accuracy ranging between 0.5 and almost 2 percentage points. We hypothesize that the strong signal of the legal entity features is caused by the fact that some legal entities have a very consistent track record because they tend to specialize in similar types of cases. On the other hand, the same cannot be said about parties, *e.g.*, a company (especially firms that specialize on opportunistic IP enforcement) may sue on many different IP issues. The latter issue does not apply to defendant parties, which explains why features extracted from defendant parties have a significant contribution.

(b) The figure shows that features extracted from judge and district entities have a significant influence on performance. This suggests that there is a significant correlation between some judges and districts and certain lawsuit outcomes (*i.e.*, plaintiff winning or defendant winning).

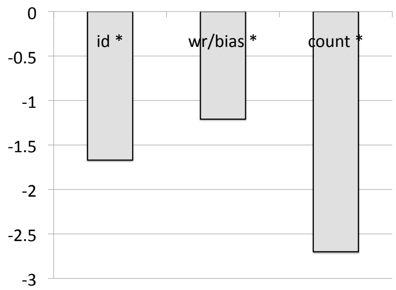

Figure 2: Ablation experiment for the CRF model, with feature groups defined around feature types (identifiers, win rates, and counts). All feature groups have a statistically-significant impact (2-tailed paired t-test, $p = 0.01$).

We performed a second ablation experiment, this time generating feature groups based on feature types instead of entity types. We followed the same procedure as in the previous experiment. The results are summarized in Figure 2. This experiment illustrates two additional important facts:

(c) Surprisingly, the `count` feature, *i.e.*, the number of past cases that had the corresponding entity as participant, is a better predictor of future behavior than the `wr` or `id` features, which capture the success rates of entities. Our conjecture is that `count` is a better estimate of experience, *i.e.*, the more cases an entity had participated in, the more experience that entity has. An additional factor in favor of the `count` feature is that the values of the `wr` feature do not account for sparsity, *i.e.*, a legal entity with few past cases may have an inflated `wr` value.

(d) The `id` feature, which models the performance of individual entities in a specific role (plaintiff or defendant), is more important than the `wr` feature, which captures entity win rates regardless of their role. This experiment shows that the historical performance of an entity in a given role can be different than its overall performance, *i.e.*, a party may be successful when litigating as a defendant but not a successful litigator overall, and it is important to represent this distinction.

5. RELATED WORK

Our work falls in the larger field of knowledge discovery from legal databases [7]. Within this field, a few works focused on predicting litigation outcomes. For example, Arditi *et al* [2] is the first work to address the problem of predicting outcomes of lawsuits. In this work, they focus on litigation in the construction industry. Their data consists of a training set of 102 construction industry related cases filed in Illinois appellate courts from 1982 through 1992 and an additional test set of 12 cases filed in the same court from 1992 through 1994. They represent each case by a set of 45 domain-specific features extracted from the case pleading documents such as party type, type of contract, contract value, etc., and train a Neural Network. This model achieved 67% accuracy on the test set. In another work, Chau applied an approach based on neural networks [4] to litigation in the same industry. The main difference is that the data came from Hong Kong courts from 1991 to 2000. This dataset had a considerably larger number of cases (1,105 overall) but the model used only 13 unique features manually extracted from the case documents. Chau trains a neural network model using particle swarm optimization, achieving an accuracy of 82%. These works make an initial step towards risk analysis using the merits of the case, even though they use either shallow approximations of case merits or features that were manually coded. On the other hand, we are the first to investigate an orthogonal direction that uses only historical or concurrent information on the litigating entities to evaluate litigation risk. Additionally, all our features are automatically extracted from the training data.

6. CONCLUSIONS

Litigation, and in particular IP litigation, is an extremely important element of the United States economy. Billions of dollars are spent each year in preparing for litigation. IP trials commonly award damages in the millions of dollars or even billions.[6] This work essentially argues (with empirical support) that IP litigation is a problem fit for forecasting. We introduced the novel problem of assessing the risk for parties involved in IP litigation based solely on prior factors. Prior factors are attractive because they capture the merits of the case indirectly, identify potential biases in the system, and are easier to extract and model than the actual merits of the case. We modeled risk estimation by estimating the probability that the corresponding party will lose the case. We built a logistic regression classifier to capture historical features and a novel relational model using conditional random fields to jointly predict the outcomes of concurrent and related cases. Our experimental results show that the CRF-based relational classifier outperforms the baseline majority classifier by more than 22 relative percentage points in accuracy. Our extensive feature analysis unearthed the entity types that are most influential in determining the risk.

Our work has established that in the IP litigation domain risk estimation systems can be developed by modeling only the prior information of the participating entities. To the best of our knowledge, this work is the first to show that this is possible. As part of future work, we would like to combine both merits of the case as well as prior factors into a single model to achieve further improvements in performance.

We believe that this work can help parties involved in an IP lawsuit make well-informed decisions in terms of settlement or continuation of a case. Having an accurate estimate of litigation risk will also reduce the number of cases that reach trial, which benefits financially all the parties involved and improves the overall efficiency of the judicial system itself.

7. REFERENCES

[1] C. Andrieu, N. De Freitas, A. Doucet, and M.I. Jordan, *An introduction to MCMC for Machine Learning*, Machine Learning, 2003.

[2] D. Arditi, F.E. Oksay, and O.B. Tokdemir, *Predicting the Outcome of Construction Litigation Using Neural Networks*, Computer-Aided Civil and Structural Engineering, 13, 1998.

[3] J. Besag, *Statistical Analysis of Non-lattice Data*, The Statistician, 24, pp. 179–195, 1975.

[4] K.W. Chau, *Prediction of Construction Litigation Outcome Using Particle Swarm Optimization*, Proc. of the International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems, 2005.

[5] J. Finkel, T. Grenager, and C. Manning, *Incorporating Non-local Information into Information Extraction Systems through Gibbs Sampling*, Proc. of the Annual Meeting of the Association for Computational Linguistics, 2005.

[6] J. Lafferty, A. McCallum, and F. Pereira, *Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data*, Proc. of International Conference on Machine Learning, 2001.

[7] A. Stranieri and J. Zeleznikow. *Knowledge Discovery from Legal Databases*, Springer, ISBN 978-1-4020-3036-9, 2005.

[6] Data provided by the Stanford IPLC project.