Towards Comprehensive Patent Approval Predictions: Beyond Traditional Document Classification

Xiaochen Gao **Zhaoyi Hou** **Yifei Ning** **Kewen Zhao** **Beilei He**
Jingbo Shang* **Vish Krishnan**
University of California, San Diego
{xig034, z9hou, y3ning, k4zhao, behe, jshang, vkrishnan}@ucsd.edu

Abstract

Predicting the approval odds of a patent application is a challenging problem involving multiple factors. The most important factor is arguably the novelty — *35 U.S. Code § 102* rejects applications that are not sufficiently differentiated from prior art. Novelty evaluation distinguishes the patent approval prediction from conventional document classification — too-similar newer submissions are considered as not novel and would receive the opposite label, thus confusing standard document classifiers (e.g., BERT). To address this issue, we propose a novel framework AISeer that unifies the document classifier with handcrafted features, particularly time-dependent novelty scores. Specifically, we formulate the novelty scores by comparing each application with millions of prior art using a hybrid of efficient filters and a neural bi-encoder. Moreover, we impose a new regularization term into the classification objective to enforce the monotonic change of approval prediction w.r.t. novelty scores, From extensive experiments on a large-scale USPTO dataset, we find that standard BERT fine-tuning can partially learn the correct relationship between novelty and approvals from inconsistent data. However, our time-dependent novelty feature and other handcrafted features offer a significant boost on top of it. Also, our monotonic regularization, while shrinking the search space, can drive the optimizer to better local optima, yielding a further small performance gain.

1 Introduction

Intellectual property (IP) is an important and integral to the economy. IP-intensive industries directly accounted for 27.9 million jobs in the U.S. (USPTO, 2016) Theoretical and empirical evidence shows that patents are effective in fostering technological progress. (Gallini, 2002; Hu and Png, 2013; Hall and Harhoff, 2012) Securing patent approvals offers a major shot in the arm to inventors and innovators, increasing the chances of obtaining angel and venture capital investments. However, the process of getting a patent approved can cost applicants tens of thousands of dollars in payments to law firms who claim to be helpful in understanding what gets approved and improving the odds of success of a patent application. Thus, algorithmic approaches to aid in the patent evaluation process can potentially save precious time and resources for applicants during the patent application phase, as well as benefit patent examiners in government patent offices around the world, accelerating and improving the review process (Ebrahim, 2018).

The approval of a *patent application*, according to U.S. patent laws, is determined necessarily and sufficiently by the approval of *application claims*. Patent laws define individual claims as the subject matter of *inventions* (*35 U.S. Code* §112), on which "patentability" is defined (*35 U.S. Code §* 101, 102, and 103) (refer to Appendix B). No overall assessment of a patent application is provisioned.

In practice, application claims demarcate the scope of legal protection that an applicant is seeking and are the eventual objects for investigation under legal disputes or transfer of commercial rights. Patent examiners from the U.S. Patent and Trademark Office (USPTO) make decisions on each application claim individually and independently with other sections as supporting materials. Therefore we focus on claim texts and use the term "*patent approval*" informally and interchangeably referring to "*claims approval*." In particular, we primarily consider *35 U.S. Code § 102*, assessing the *novelty* of application claims.

To the best of our knowledge, we are the first to try to predict patent (claim) approval, which is as an extremely challenging problem for multiple reasons. First, patent documents comprise of technically nuanced and challenging to parse language (intricate legalese). Patent texts are usually legal and technical descriptions of objects or pro-

* Jingbo Shang is the corresponding author.

Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics
Volume 1: Long Papers, pages 349 - 372

cesses, which tend to be complex in vocabulary and grammatical structures (Singer and Smith, 1967). Claims are examined not only literally, but also for their legal implications. Appendix A provides a few example application claims.

Second, the patent examination process tends to suffer from subjectivity and inconsistencies (O'Neill, 2018a), exemplified by variance across offices and groups, (O'Neill, 2018b) and across human examiners. In FY17, only 66% of primary examiners are within a 12.5% delta off the average allowance rate (USPTO, 2017).

Third, at the core of patent examination, evaluation of novelty is time-dependent. Rejections of claims by *35 U.S. Code § 102* require examiners to *cite* prior approved patent claims, *prior art*, as evidence. More details about the examination process can be found in USPTO (2020). The United States Patent and Trademark Office (USPTO) receives thousands of applications a week; thus a novel application at one time may be dramatically different in the assessment of novelty after a short time period. This means that a classifier can pick up a positive label from an earlier approved application but receive a negative label from an application sometime later with similar technical content, which is deemed no longer novel. Such conflicting information can confuse the classifier and undermine its performance. In other words, the data labels are intrinsically noisy and inconsistent due to the nature of the domain problem.

Although AI/ML approaches are often discussed in the patent domain (Aristodemou and Tietze, 2018) such as in the area of information retrieval (Kang et al., 2007; Fujii, 2007; Shalaby and Zadrozny, 2019), applications of deep NLP methods are mostly concerned with classifying the content domains of patents (Verberne et al., 2010; D'hondt et al., 2013; Hu et al., 2016; Lee and Hsiang, 2019). In addition, the extant literature usually explores approved patents rather than applications (Balsmeier et al., 2018). Even to simply classify the topics of approved patents, state-of-the-art document classifiers can only achieve an accuracy of about 69.3% - only 2.2% over RoBerta (Zaheer et al., 2020). Due to these issues, patent approval prediction task is much more challenging than topic classification for document classifiers.

To mitigate the issues, we first develop several handcrafted features based on domain knowledge for use alongside the language model for context

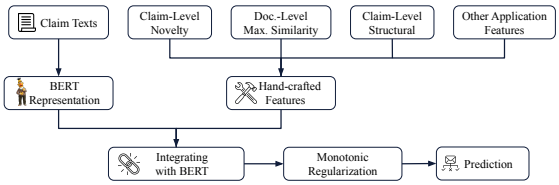


Figure 1: An overview of our proposed AISeer.

and control. The time-dependent nature of the novelty also makes traditional document classifiers not suitable here, because they typically assume that similar instances belong to the same label. To address this challenge, we propose a novel framework AISeer as shown in Figure 1. We formulate a time-dependent novelty score for each patent claim with its semantic similarity against prior approved claims from *patent grants*, which are final versions of *approved* patents. Specifically, inside a comprehensive pool comprising millions of grants, we consider those approved before the filing date of the focal application and then measure the maximum semantic similarity score of the focal patent claim matched with all approved claims in the time-dependent sub-pool. To improve computing efficiency, we apply document-level filters to narrow the sub-pool for each claim. Integrating such similarity scores with handcrafted features and BERT, we conduct experiments on the large-scale USPTO dataset and find significant performance gains over fine-tuning a standard BERT alone.

All else equal, a patent claim with a higher similarity score, i.e., semantically more similar to prior approved claims, should be less likely to be approved. Hence we propose to impose monotonic regularization on the novelty score so that the loss function has an additional term of the hinge loss to further penalize non-decreasing predictions in the similarity. This effectively restricts the search space for the optimizer to prediction mechanisms that are reasonably consistent with the novelty measure. From our experiments, this regularization significantly impacts the model outputs. Although performance improvements are limited, it can help the optimizer steer away from unfavorable local optima and further improve AUROC. We further discuss the experimental findings in depth to illustrate how BERT and handcrafted features contribute in overcoming the unconventional data issues.

In summary, our contributions are as follows.

- We collect patent application data from several data sections of USPTO and integrate full texts,

Table 1: Dataset Statistics. The approval ratio is calculated based on *35 U.S. Code § 102* labels.

	Train	Validation	Test
Applications M	216,101	175,597	153,632
Claims N	3.90M	3.07M	2.58M
Approval %	80.65	80.16	81.68
Time range	04/16-02/17	03/17-10/17	11/17-06/19

metadata, office actions, rejections and citations data into a massive dataset;

- We develop a series of handcrafted features to aid the prediction of *35 U.S. Code § 102* approval decisions. In particular, we design and analyze a time-dependent feature that measures the novelty of patent applications at the time of filing;
- We incorporate the handcrafted features and impose monotonic regularization on the novelty features to shed light on how the intrinsic data inconsistency issues in the domain problem can be mitigated.

Reproducibility. We will release the benchmark dataset and our code on GitHub: https://github.com/acl-2022-towards-comprehensive/acl-2022-camera-ready.

2 Problem Formulation and Benchmark

In this section, we formally formulate the novelty-based patent approval problem. We describe the experiment setup, the dataset, and baseline results with common document classifiers.

2.1 Problem Definition and Formulation

We follow legal definitions under *35 U.S. Code § 102*. Despite the popular notions of patent approval or issues, what is actually being approved/rejected are individual claims.

Each patent applications A_k, $k \in \{1 \cdots M\}$, sorted by filing dates, comprises of a number of application claims. Given text representation $\mathbf{X}_i$, $i \in \{1 \cdots N\}$, of each application claim, there exist $\{i_k\}$, $k \in \{0 \cdots M\}$ such that claim representations $\{\mathbf{X}_{i_{k-1}} \cdots \mathbf{X}_{i_k}\}$ belong to patent application A_k. Binary labels y_i indicate approval decisions derived from patent rejections and office actions data where $y_i = 1$ indicates claim approvals. We would like to classify application claims according to approval labels.

2.2 Benchmark Dataset Preparation

Dataset Collection. USPTO provides public data arranged in separate sources, including application and grant full texts, application metadata, citations, office actions, and rejections (USPTO, d,e). Patent grants are final versions of approved patent applications. Later we will utilize grants for constructing the application novelty feature. To extract labels and create handcrafted features, we utilize both the legacy data system for office actions, rejections and citations made between 2008 and mid-2017 (Lu et al., 2017), and newer v2 APIs that cover mid-2018 onward. For application metadata, we obtain bulk data from PEDS (Patent Examination Data System) (USPTO, b). In order to match all the available labels, we obtain weekly bulk releases for of both utility patent applications and utility patent grants in XML format ranging between 2005 and 2019. In total, we extract 8.8 million patent applications and 3.7 million patent grants during the same time period whose texts are around 730 GB.

Dataset Processing. According to patent laws, only one version among possibly a number of revisions is published and available as full-text data. Meanwhile, for a considerate amount of applications, the entire history of office action data and rejection data are available, where allowances or rejections for each individual claim under each legal clause are formally made. Hence we ought to identify the labels associated with the published version among patent examination rejection data and office action data.

We take a "snapshot" approach. Given the available publication version of each application as snapshots, the examination decisions of each claim particularly with respect to the snapshot version are processed and attached as classification labels. Therefore, with the huge number of snapshots, regardless of the subsequent actions of the applicant, e.g. abandonment, the model can be kept agnostic of the status in the application pipeline. This way, we allow the model to predict for any version of a patent application so that the attorneys and applicants can evaluate their chances for decision making. Technical preparations for publication of an application generally begin 4 months prior to the projected date of publication. Hence we match the closest office action dates with publication dates minus 4 months which is supposed to be the benchmark date for the available version, so that correct labels can be obtained. Please refer to the essential publication regulations in Appendix D.

Data are merged by the application number and ingested into a DBMS. We find out around 900K ap-

plications under which all corresponding sections of data are available. Because of the data size and to control for computation times, we choose the most recent, around 500K applications and effectively around 9.5 million claims for experiments.

Dataset Splits. We split the data into training, validation, and testing sets by their filing dates. The more recent ones are chosen for testing. The size for final experimental data, including the abstract, claim texts, labels, and handcrafted features, is around 15 GB. The dataset is highly imbalanced towards positive labels (see Table 1 Approval %).

2.3 Common Document Classifier Benchmark

Common Document Classifiers. We mainly evaluate the following common document classifiers.

- **Log. Reg.** refers to logistics regression using TF-TDF features.
- **Text-CNN** (Kim, 2014) with GloVe (Pennington et al., 2014) embeddings as the input. Adam optimizer with learning rate 0.001. 10 epochs' run; batch size as 1024;
- **LSTM** (Hochreiter and Schmidhuber, 1997) with GloVe embeddings as the input. AdamW optimizer with learning rate 0.005 and 10 epochs' run; batch size as 1024;
- **BERT** (Devlin et al., 2018) fine-tuning. AdamW optimizer with learning rate 5e-5 as the optimizer. The number of fine-tuning epochs as 5; batch size as 256. This is the the same model as in the state-of-the-art model, **PatentBERT**, in patent content classification (Lee and Hsiang, 2019) with a different set of hyper-parameters and balanced class weights. The original PatentBERT model is designed for a different task, and the experimental setting is not suitable for predicting patent approvals, hence we make the tweaks.

In all of the models, we impose class weights in the loss functions inversely proportional to the number of class instances, such that two classes are treated equally by the optimizer. For the details, please refer to Section 3.1. The neural models are trained with text inputs processed at a maximum length of 128 tokens per claim and on a single GPU.

Evaluation Metrics. Given the imbalanced nature of our dataset, we adopt both the Area Under the Curve for the ROC plot (Fawcett, 2004) (AUROC) and macro F1 score as our evaluation metrics. With AUROC, the predicting performance of the minority class could be taken into consideration with a similar weight as for the majority class (in our case, positive class). Moreover, the probability-based metric can provide more detailed insights into model performances. Therefore, we choose **AUROC** as our **main metric**. The **macro F1** score is a direct average of F1 scores of both the positive class and the negative class and provides an alternative balanced view of both classes' performances. We treat it as a **secondary metric**. We compute the maximum macro F1 score (Lipton et al., 2014) by varying the decision threshold for each model. Other traditional measures focused on the positive class performance such as accuracy and recall have little practical implications due to data imbalance.

Table 2: Benchmarking Common Document Classifiers.

	AUROC %	**Macro F1 %**
Random Guess	50.00	50.00
Predicting All "1"	50.00	44.96
Log. Reg. (Tf-Idf)	58.94	54.54
TextCNN (GloVe)	59.70	55.58
LSTM (GloVe)	61.68	**56.95**
BERT (PatentBERT)	**61.79**	56.51

Benchmark Results. Table 2 shows common document classifiers' performance with some naive predictions as references. Results of neural models are reported with the median metrics among several runs with different optimizer random states. Figure 4 in Appendix F further visualizes more details of the ROC curves of these models. One can find that BERT and LSTM are arguably the most effective ones. Therefore, we will focus on BERT and LSTM for further comparisons.

3 The AISeer Framework

Our AISeer framework unifies the document classifier, handcrafted features and monotonic regularization, as shown in Figure 1. It is compatible with almost all document classifiers. In this paper, we choose BERT as the base document classifier to demonstrate the effects as it is widely adopted and also performs well in our benchmark evaluations. After each application claim text is run through the BERT model, the output representation is concatenated with the corresponding handcrafted features. Our handcrafted features include a time-dependent claim-level novelty score, claim-level structural features, document-level similarity scores, and other application metadata features. We further impose a monotonic regularization on the impact of the claim-level novelty score so that the loss function has an additional term of the hinge loss.

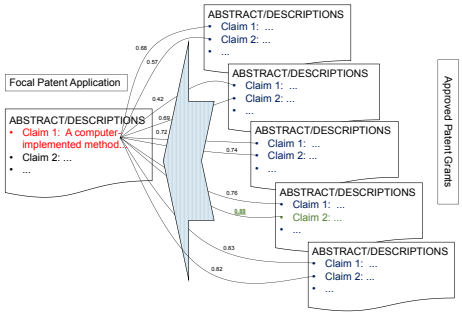


Figure 2: Illustration of Novelty Feature Construction.

3.1 Base Document Classifier

For self-containedess, we briefly introduce how we use BERT in AISeer. We first utilize BERT to transform the i-th application claim to a text representation $\mathbf{X_i}$ in batches of a size N_b, which is then passed to a linear layer to obtain the prediction through a softmax layer.

Approvals (i.e., $y_i = 1$) are much more popular than rejections (refer to Table 1), so the vanilla training will bias the model towards approvals. Therefore, we adopt a weighted loss for training: $\mathcal{L} = \sum_i -w_{y_i}\left(y_i \log \hat{y}_i + (1-y_i)\log(1-\hat{y}_i)\right)$ where w_{y_i} denotes the fixed weights of the two classes, which is inversely proportional to the number of instances from the corresponding class, balancing the training weights of the two classes.

3.2 Claim-Level Novelty Feature $\mathbf{N_{s,claim}}$

The backbone of the novelty feature is the *time-dependent* claim-level maximum similarity score.

We first index all patent grants with ElasticSearch (NV). Given a patent application and a claim under it, we first take advantage of its fast BM25-based document-level fuzzy matches to obtain the 5 most similar grant documents to the focal application document as a first-stage pre-filter. To account for time-dependence, each focal application is matched against a sub-pool of patent grants which are time-stamped to be approved strictly before the filing date of the focal application. In application level matching, all document sections are considered, including the abstract, summary of invention, and details of invention of all claims.

Among all claims under the top-5 matched grants, we then find the most similar one to the focal claim using sentence-transformer (Reimers and Gurevych, 2019) with `stsb-roberta-large` pre-trained bi-encoder model. Base cross-encoder transformers such as BERT can lack in performance for pure semantic similarity tasks. Although certain cross-encoders have excellent semantic similarity performance, it can be computationally too demanding for our purpose since the scale of the claims in all patent grants is more than 100 million, and since each grant claim can be required to be paired many times with a focal application claim. The Elasticsearch-based pre-filter process also helps manage the computational need.

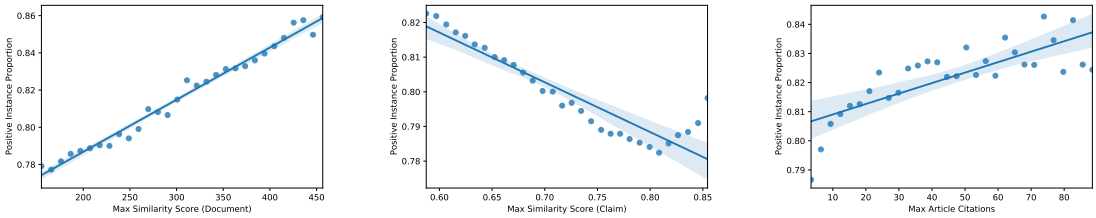
Figure 2 demonstrates how the time-dependent novelty feature is generated — the application that the red-highlighted focal claim belongs to is first matched with 5 patent grants on the application level; then the focal claim is matched against every claim under the 5 matched grants to compute the semantic similarity score, before the most similar grant claim is identified. Our experiments confirm that the claim-level maximum similarity score, as expected, is negatively correlated with *35 U.S. Code § 102* labels, as shown in Figure 3.

3.3 Application-Level Handcrafted Features

Application-Level Similarity. We consider the application-level maximum similarity score, denoted as $N_{s,doc}$, and mean similarity score generated by ElasticSearch (NV) as handcrafted features. These document-level scores measure how similar overall are the applications to the approved grants. The document-level similarity scores are positively correlated with *35 U.S. Code § 102* labels. We believe that they primarily capture the overall writing quality and the common language patterns of approvable applications.

Features from Metadata. The USPTO dataset offers a rich collection of metadata about each patent application. We use the following two of them:

- *Patent Classification*: the USPC class designated for the applications. USPC (USPTO, a) is a system of classifying the subject matter of each patent application for recording, publication, and assignment purposes. Different classes of patents tend to have varying approval rates (see Table 6 in Appendix C).
- *Number of Applicant Cited References*: the number of citations of other patents or articles initiated by the applicant herself. In the patent domain, most *citations* are initiated by the examiners as "prior arts" to reject application claims. However, they can also be made by the applicant to demonstrate understanding of related work and claim contributions. The number of applicant-initiated citations is a signal of the effort and



(a) $N_{s,doc}$ (b) $N_{s,claim}$ (c) Max Article Citation

Figure 3: Handcrafted Features vs. Proportions of Positive 102-Labels. Features are grouped into bins for 10-90 percentile against mean positive label proportions.

research the applicant puts in the application.

Other Application-Level Features are also considered for utility and writing as follows.

- *Max Citation*: based on ElasticSearch pre-filter, the maximum number of total citations among the top 5 most similar patent grant documents to the focal patent application.
- *Max Article Citation* refers to the maximum number of citations which are research articles (not other patents) in top matched grants.
- *Lexical Diversity*: the richness in the vocabulary of the abstract of the patent application.

3.4 Claim-Level Structural Features

We consider two indicators for each claim.

Component refers to indicator on whether the application claim is describing the components of a system (e.g., a machine, a process, a compound). Other claims may describe the properties or utility of particular components. This is identifiable by the transitional phrases used in the claim.

Transitional Phrase refers to indicator on whether a component claim is *open*, *closed*, or *half-open*, which is determined by which transitional phrase is used. Openness or closedness regulates the scope of legal IP protection the applicant enjoys once the patent is approved. Often it is a strategic choice by the applicant and the attorney. If a claim is *open*, indicated by transitional phrases "comprising" and legal synonyms, any additional components later added to the system are also protected, in contrast to closed claims . Open claims are more difficult to be approved. Other examples of transitional phrases include "consisting essentially of" and "consisting of". These particular language phenomena are well-known in the IP communities and sometimes referred to as "patentese" (Singer and Smith, 1967). The patent examination manual explicitly discusses these phrases with case law (USPTO, c; Silverman and Stacey, 1996).

3.5 Integrating with BERT

Now let $\mathbf{H_i}$ denote other handcrafted features in addition to $N_{s,claim}$ and $N_{s,doc}$. Figure 3 demonstrates the correlations between some representative handcrafted features and the positive label. Let $\mathbf{Z}_i = \mathbf{X_i} \cup \mathbf{H_i} \cup N_{s,claim} \cup N_{s,doc} \cup \{1\}$, $\forall i \in \{1, ..., N_b\}$. Note that X_i is the representation for the claim and that the document or application-level handcrafted features will be augmented to each claim. The concatenated Z_i will pass through the linear and the softmax layer.

3.6 Monotonic Regularization

Mathematically, we restrict the search space upon $N_{s,claim}$, regularizing predictions to be decreasing in it. The optimizer will potentially be able to find alternative paths to avoid undesirable local minima. Let $\tilde{\mathbf{Z}}_\mathbf{i}$ denote all other inputs except $N_{s,claim}$.

We manipulate the input such that inconsistency with the monotonicity in $N_{s,claim}$ is represented. The novelty scores need to be manipulated and multiplied by For a positive constant $C(0 < C < 1)$ let $N'_{s,claim} = CN_{s,claim}$, let $\mathbf{Z'_i} = \tilde{\mathbf{Z}}_\mathbf{i} \cup N'_{s,claim}$. Applying such a manual constraint on the input novelty representation completes such a monotonically decreasing relationship between input $N_{i,k}$ and output Given log-likelihood with respect to $\mathbf{Z_i}$,

$$F(\mathbf{Z_i}) = y_i \log \hat{y}_i(\mathbf{Z_i}) + (1 - y_i) \log(1 - \hat{y}_i(\mathbf{Z_i})),$$

we shall constrain $F(\mathbf{Z_i}) < F(\mathbf{Z'_i})$. To implement it, we shall impose a hinge loss penalty whenever $F(\mathbf{Z_i}) > F(\mathbf{Z'_i})$ and return 0 when otherwise.

Therefore, the final objective function becomes:

$$\mathcal{O} = \mathcal{L} + \lambda \sum_i \max\left\{0, F(\mathbf{Z_i}) > F(\mathbf{Z'_i})\right\},$$

where λ determines the regularization strength.

4 Experiments

In the experiments, we seek to answer a number of questions. To begin with, we evaluate how hand-

Table 3: Evaluation Results of AISeer, Compared Methods, and Ablations.

	AUROC%	Macro F1%
LSTM (GloVe)	61.68	56.95
BERT (patentBERT)	61.79	56.51
AISeer	**64.14**	**57.92**
Log. Reg. Feat. Only	60.45	55.47
AISeer w/o Regu.	63.71	57.73

crafted features can help the deep language model, i.e. BERT, adapt to a complex domain that differs from typical NLP use cases. In particular we focus on the novelty feature critical to patent approvals. We are interested in the extent to which a standard BERT application can learn from highly noisy labels and inconsistent data and find out the novelty pattern, i.e. the significance of novelty in determining patent approval outcomes. In addition, we study if the combination of BERT and handcrafted features serves to be adequate in capturing the novelty pattern. We also examine if monotonic regularization boosts the learning process to further overcome the intrinsic data inconsistencies.

We mainly compare **AISeer** with two baseline models, **BERT** and **LSTM**, as they are the best common document classifiers from our benchmark results. For ablation study purpose, we also compare with **Log. Reg. Feat. Only**, a logistics regression model with handcrafted features only, and **AISeer w/o Regu.**, which is a BERT model integrated with our handcrafted features but not regularized by our monotonic constraints. AISeer is trained with the same set of hyper-parameters as BERT: maximum token length as 128, fine-tuning for 5 epochs; batch size as 256; AdamW with learning rate being 5e-5 as the optimizer. The monotonic regularization parameter C is $\frac{1}{2}$ and λ is 5e-4. The models are trained on a single Nvidia Quadro RTX 8000 GPU.

The results are shown below in Table 3. The reported numbers are median results from 3 runs under the same hyperparameter setup.

4.1 Overall AISeer Results

The baseline BERT model gives decent AUC (ROC) and macro F1. The full-fledged AISeer, combining handcrafted novelty feature along with other computed ones and motonic regularization, helps with both the metric dimensions: AISeer boosts AUROC by around 2.5% percent and macro F1 by around 1% compared to the best common document classifiers. Figure 5 in Appendix F shows the AUROC improvement originates consistently from the entire spectrum of prediction scores.

Aforementioned in the introduction, when simply classifying the topics of approved patents, state-of-the-art document classifiers can only achieve an accuracy of about 69.3% (only 2.2% over RoBerta) (Zaheer et al., 2020). Given the difficulty level and subjective nature of the patent approval task, the performance improvement is non-trivial and practically impactful. Standard BERT fine-tuning realizes an AUROC increase of only 11.79% over completely random or naive predictions, which also exemplifies the problem's difficulty. Our approach achieves an additional performance of 2.35%, which is equivalent to 20% of the total benefits of the original BERT model. Given that BERT remains one of the most effective models in varieties of NLP tasks, and especially that handcrafted features have relatively low dimensionality compared to BERT, we believe that the performance gain equivalent to 20% of the performance gain of BERT is substantial for this completely new application domain.

The lower half of Table 3 shows the result of Log. Reg. Feat. Only, indicating the necessity of a language model. Neither a language model only nor handcrafted features only can yield satisfactory performance.

4.2 Evaluating Handcrafted Features

Comparing AISeer w/o Regu. result, also in the lower half of Table 3, and the standard BERT and LSTM results, it is shown that handcrafted features improve on best common document classifiers by about 2%. We believe that the handcrafted features combined, in particular, the novelty feature, helps in resolving label contradictions and data inconsistency.

We believe the novelty feature should be only considered under contexts and will not perform well on its own. First, novelty can be a subjective concept and may vary according to different types of claims, openness of claims, the department (category), etc. Second, novelty as practically measured by dis-similarity, can be easily achieved by poorly written random content, thus structural or overall similarity is also important. However, the observations indicate that there are potential conflicts between the novelty feature and other handcrafted features. While the latter helps with prediction performance on their own and provide contexts for the

novelty feature thus imperative, it will also attenuate the effects of the regularized novelty feature. We leave this challenge for future work.

One may also ask whether the handcrafted features have contributed significantly given the moderate improvement. Granted, application full texts may also contain signals for the patent class and applicant efforts that may partially reflect handcrafted features and the document classifier such as BERT may pick them up.

To shed light on how AISeer learns from handcrafted features, we run linear regressions for the model prediction scores on handcrafted features for interpretable insights and present statistical results, as shown in Table 4. In the table, even prediction scores under BERT are significant in all handcrafted features, showing that BERT does learn knowledge overlapping with the handcrafted features to some extent.Overall, low R^2's indicate that knowledge from the deep neural model and knowledge from handcrafted features are quite distinct.

Comparing BERT and AISeer w/o. Regu., the significant R^2 increase from 0.085 to 0.125 shows that AISeer captures handcrafted features much more effectively than BERT. The prediction scores of AISeer w/o. Regu. have an additional about 4% increase in explanability by the handcrafted features.

4.3 Evaluating BERT Learning and Monotonic Regularization

In Table 3, comparing AISeer and AISeer w/o Regu., the median run result indicates that adding monotonic regularization produces a small magnitude of improvement. Table 4 also provides insights with respect to the monotonic regularization. According to Table 4, our claim-level novelty feature $N_{s,claim}$ has the most significant impact, i.e. the coefficients are much larger in every column. The use of monotonic regularization alone boosts the R^2 significantly, indicating that the approach also helps the model learn from handcrafted features overall.

About 19% of the knowledge of AISeer corresponds to handcrafted features, a 10% increase over BERT. Also, AISeer corrects incorrect coefficient signs from BERT. Intuitively, the approval chance shall increase with the number of applicant cited references. However, BERT prediction scores are negatively correlated with it statistically significantly. Under AISeer, this direction is reversed to match intuitions.

We also evaluate the Spearman correlation coefficients of the probability prediction scores produced by the models with the claim-level novelty feature Pearson correlations with the document-level similarity score. Spearman correlations measure the strength and direction of monotonic association between two variables. According to Table 5, first we can confirm that applying monotonic regularization significantly pushes the prediction scores to be more monotonically decreasing in the core novelty feature – the Spearman correlation shifts from -0.0230 to -0.103. However, compared to the BERT, the regularization effect is less prominent. Observe that adding handcrafted features will actually steer the monotonicity into the opposite direction. Our regularized AISeer model manages to both benefit from the novelty feature and incorporate knowledge from other handcrafted features.

While Table 4 illustrates the significant effects of applying the monotonic regularization on the prediction scores, we acknowledge that the observed main performance improvement is not very significant. In fact, although monotonic regularization raise the performance on average, it does not always yield desirable improvements depending on the random seed and the hyperparameter setup.

The BERT model may already have a decent learning power to mine the novelty measurement despite the noisy data. We observe that in Table 4, the BERT prediction scores, learned from texts only, are significant in the novelty feature and are in the correct direction. The relatively small performance gain of using monotonic regularization may also be attributed to the compromised precision of the novelty feature due to the use of the ElasticSearch pre-filter for the sake of computational costs.

5 Related Work

To our knowledge, our work is the first in predicting patent approvals according to the examination procedures at the government patent office. Few extant researches attempt to predict decisions in office. Winer (2017) studies PTAB (Patent Trial and Appeal Board) hearing decisions at USPTO. Other related work addresses patent quality in a general and broad sense (Wu et al., 2016). More broadly in the IP/patent domain, although AI/ML applications have been often advocated (Ebrahim, 2018), studied (for a review see (Aristodemou and

Table 4: Regression Analysis of Prediction Scores on Handcrafted Features.

	BERT	AISeer w/o Regu.	AISeer
No. of Applicant Cited Refs	-3.5e-06*** (9e-7)	-8.2e-06*** (1e-6)	4.3e-06*** (8e-7)
Transitional Phrase - Open	-0.045*** (0.000)	-0.037*** (0.000)	-0.067*** (0.000)
Transitional Phrase - Closed	-0.015*** (0.000)	-0.022*** (0.000)	2e-4 (0.000)
Max Article Citations	1.9e-5*** (7e-7)	2.5e-5*** (7e-7)	3.2e-5*** (5e-7)
$N_{s,doc}$	2e-4*** (6e-7)	4e-4*** (6e-7)	2e-4*** (4e-7)
$N_{s,claim}$	-0.18*** (0.001)	-0.17*** (0.001)	-0.21*** (0.001)
$\mathbf{R^2}$	0.085	0.125	0.189

Notes: HC1 heteroskedasticity-robust standard errors used. Not all regressors shown. ***1% significance level.

Table 5: Correlations between Features and Predictions.

	BERT	AISeer w/o Regu.	AISeer
$N_{s,doc}$ (Pearson)	0.128	0.238	0.180
$N_{s,claim}$ (Spearman)	-0.0788	-0.0230	-0.103

Tietze, 2018)) or implemented in practice (Lu et al., 2017), most work focus on determining patent content classes to save manpower or concern only with patent grants rather than applications (Verberne et al., 2010; D'hondt et al., 2013; Hu et al., 2016; Balsmeier et al., 2018; Lee and Hsiang, 2019). Recent studies (Hsu et al., 2020) emerge aiming at predicting patent transfers and the economic value.

Other streams of related work include those exploring patent similarity. Our approach of constructing the novelty feature with a state-of-the-art neural bi-encoder (Reimers and Gurevych, 2019) is significantly more advanced than relatively rudimentary approaches in the extant literature, such as text matching and frequency-based methods (Younge and Kuhn, 2016; Arts et al., 2018; Shahmirzadi et al., 2019). Studies on semantic analysis and representation of technology (Kim et al., 2016; Strumsky and Lobo, 2015) based on patent data are also related.

6 Conclusions and Future Work

In this paper, we tackle the challenging problem of predicting patent approval decisions as per *35 U.S. Code § 102*, namely the novelty-based decisions. We have prepared a large-scale benchmark dataset by consolidating different data sources from USPTO. From the evaluations of the popular document classifiers, BERT and LSTM are arguably the most effective ones. We identify the time-dependent challenge of the novelty judgement, and therefore propose AISeer, a novel framework going beyond the traditional document classifiers. Specifically, we construct a claim-level core novelty feature along with several other handcrafted features and apply them on top of the pre-trained BERT model. We further propose to add the monotonic regularization on the core novelty feature to resolve the potential label conflicts caused by the mechanism of the patent examination process. Experimental results have verified the superiority of AISeer and also the effectiveness of introducing novelty features and monotonic regularization.

We believe that our work is beneficial to various parties, including patent applicants, attorneys, examiners and regulators. While the advantages of our regularization methodology are significant, there is still room for potential metric improvements, thus further developing the work will yield opportunities for promising future research and greater contributions to the communities. In future, it is important to extend the scope from claims to the other sections in the patent applications. Relationships among components and entities described in claims and relations among claims are also critical to investigate.

Acknowledgments

We want to thank the anonymous reviewers for their insightful comments. The research was sponsored in part by National Science Foundation Convergence Accelerator under award OIA-2040727 as well as generous gifts from Google, Adobe, and Teradata. Any opinions, findings, conclusions, or recommendations expressed herein are those of the authors and should not be interpreted as necessarily representing the views, either expressed or implied, of the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for government purposes notwithstanding any copyright annotation hereon.

Ethical Consideration

This paper focuses on the patent approval prediction. The data is publicly available from USPTO and we collected it to form a large-scale dataset. Our architecture is built upon open-source models and all the datasets are available online. Therefore, we do not anticipate any major ethical concerns.

References

Leonidas Aristodemou and Frank Tietze. 2018. The state-of-the-art on intellectual property analytics (IPA): A literature review on artificial intelligence, machine learning and deep learning methods for analysing intellectual property (IP) data. *World Patent Information*, 55:37–51.

Sam Arts, Bruno Cassiman, and Juan Carlos Gomez. 2018. Text matching to measure patent similarity. *Strategic Management Journal*, 39(1):62–84.

Benjamin Balsmeier, Mohamad Assaf, Tyler Chesebro, Gabe Fierro, Kevin Johnson, Scott Johnson, Guan-Cheng Li, Sonja Lück, Doug O'Reagan, Bill Yeh, et al. 2018. Machine learning and natural language processing on the patent corpus: Data, tools, and new measures. *Journal of Economics & Management Strategy*, 27(3):535–553.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Eva D'hondt, Suzan Verberne, Cornelis Koster, and Lou Boves. 2013. Text representations for patent classification. *Computational Linguistics*, 39(3):755–775.

Tabrez Y Ebrahim. 2018. Automation & predictive analytics in patent prosecution: USPTO implications & policy. *Ga. St. UL Rev.*, 35:1185.

Tom Fawcett. 2004. ROC graphs: Notes and practical considerations for researchers.

Atsushi Fujii. 2007. Enhancing patent retrieval by citation analysis. In *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 793–794.

Nancy T Gallini. 2002. The economics of patents: Lessons from recent US patent reform. *Journal of Economic Perspectives*, 16(2):131–154.

Bronwyn H Hall and Dietmar Harhoff. 2012. Recent research on the economics of patents. *Annu. Rev. Econ.*, 4(1):541–565.

Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation*, 9(8):1735–1780.

Po-Hsuan Hsu, Dokyun Lee, Prasanna Tambe, and David H Hsu. 2020. Deep learning, text, and patent valuation. *Text, and Patent Valuation (November 16, 2020)*.

Albert GZ Hu and Ivan PL Png. 2013. Patent rights and economic growth: Evidence from cross-country panels of manufacturing industries. *Oxford Economic Papers*, 65(3):675–698.

Mengke Hu, David Cinciruk, and John MacLaren Walsh. 2016. Improving automated patent claim parsing: Dataset, system, and experiments. *arXiv preprint arXiv:1605.01744*.

In-Su Kang, Seung-Hoon Na, Jungi Kim, and Jong-Hyeok Lee. 2007. Cluster-based patent retrieval. *Information processing & management*, 43(5):1173–1182.

Daniel Kim, Daniel Burkhardt Cerigo, Hawoong Jeong, and Hyejin Youn. 2016. Technological novelty profile and invention's future impact. *EPJ Data Science*, 5(1):1–15.

Yoon Kim. 2014. Convolutional neural networks for sentence classification. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1746–1751, Doha, Qatar. Association for Computational Linguistics.

Jieh-Sheng Lee and Jieh Hsiang. 2019. Patentbert: Patent classification with fine-tuning a pre-trained bert model. *arXiv preprint arXiv:1906.02124*.

Zachary C Lipton, Charles Elkan, and Balakrishnan Naryanaswamy. 2014. Optimal thresholding of classifiers to maximize F1 measure. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 225–239. Springer.

Qiang Lu, Amanda Myers, and Scott Beliveau. 2017. USPTO patent prosecution research data: Unlocking office action traits.

Elastic NV. The heart of the free and open elastic stack.

Jeff O'Neill. 2018a. Visualizing outcome inconsistency at the USPTO. *IPWatchdog.com*.

Jeff O'Neill. 2018b. Visualizing outcome inconsistency by group at the USPTO. *IPWatchdog.com*.

Jeffrey Pennington, Richard Socher, and Christopher D Manning. 2014. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1532–1543.

Nils Reimers and Iryna Gurevych. 2019. Sentence-BERT: Sentence embeddings using siamese bert-networks.

Omid Shahmirzadi, Adam Lugowski, and Kenneth Younge. 2019. Text similarity in vector space models: a comparative study. In *2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA)*, pages 659–666. IEEE.

Walid Shalaby and Wlodek Zadrozny. 2019. Patent retrieval: a literature review. *Knowledge and Information Systems*, pages 1–30.

Arnold B. Silverman and George K. Stacey. 1996. Understanding" patentese"—a patent glossary. *JOM*, 48(9):77–79.

TER Singer and Julian F Smith. 1967. Patentese: A dialect of english? *Journal of Chemical Education*, 44(2):111.

Deborah Strumsky and José Lobo. 2015. Identifying the sources of technological novelty in the process of invention. *Research Policy*, 44(8):1445–1461.

USPTO. a. Index to the United States Patent Classification (USPC) system. *uspto.gov*.

USPTO. b. Patent Examination Data System. *uspto.gov*.

USPTO. c. Patent Glossary. *uspto.gov*.

USPTO. d. USPTO APIs. *uspto.gov*.

USPTO. e. USPTO Datasets. *uspto.gov*.

USPTO. 2016. Intellectual property and the U.S. economy. *uspto.gov*.

USPTO. 2017. Consistency in decision making: Differences in allowance rates among similarly situated primary examiners. *uspto.gov*.

USPTO. 2020. Manual of Patent Examining Procedure (MPEP) Ninth Edition. *uspto.gov*.

Suzan Verberne, EKL D'hondt, NHJ Oostdijk, and Cornelis HA Koster. 2010. Quantifying the challenges in parsing patent claims.

David Winer. 2017. Predicting bad patents: Employing machine learning to predict post-grant review outcomes for US patents.

Jheng-Long Wu, Pei-Chann Chang, Cheng-Chin Tsao, and Chin-Yuan Fan. 2016. A patent quality analysis and classification system using self-organizing maps with support vector machine. *Applied soft computing*, 41:305–316.

Kenneth A Younge and Jeffrey M Kuhn. 2016. Patent-to-patent similarity: A vector space model. *Available at SSRN 2709238*.

Manzil Zaheer, Guru Guruganesh, Kumar Avinava Dubey, Joshua Ainslie, Chris Alberti, Santiago Ontanon, Philip Pham, Anirudh Ravula, Qifan Wang, Li Yang, et al. 2020. Big bird: Transformers for longer sequences. In *NeurIPS*.

A Example Patent Claims

Example 1: *A computer-implemented method for managing deep-learning, the method comprising: deploying a first and a second scoring endpoint with payload logging for a deep-learning model; receiving, at the second scoring endpoint, native data and a user-generated score for the native data; pre-processing, at the second scoring endpoint, the native data into readable data for the deep-learning model; outputting, from the second scoring endpoint to the first scoring endpoint, the user-generated score for the native data and the readable data, wherein the first scoring endpoint is associated directly with the deep-learning model; outputting, from the second scoring endpoint to a payload store, a raw payload, wherein the raw payload includes the native data; processing, at the first scoring endpoint and using the deep-learning model, the readable data and the user-generated score to output a transformed payload and a prediction, respectively, to the payload store; matching, at the payload store, the raw payload with the transformed payload and the prediction to produce a comprehensive data set; evaluating the comprehensive data set to describe a set of transformation parameters; and retraining the deep-learning model to account for the set of transformation parameters.*

Example 2: *A system for use in allowing a user to conduct one or more transactions at one or more touchpoints in a business facility, the system comprising: an authentication component configured to authenticate the user as a person allowed to conduct the one or more transactions; a tracking component configured to track the user's location within the facility as the user moves through the facility; and a control component configured to: receive authentication information from the authentication component; receive location information from the tracking component; use the location information to recognize that the user has moved into position to engage one of the touchpoints; and deliver a message to the touchpoint authorizing the touchpoint to engage in one or more transactions with the user.*

Example 3: *A hybrid nano-filament composition for use in a lithium battery cathode, said composition comprising: a) An aggregate of nanometer-scaled, electrically conductive filaments that are substantially interconnected, intersected, or percolated to form a porous, electrically conductive filament network, wherein said filaments have a length and a diameter or thickness with said diameter or thickness being less than 500 nm; and b) Micron- or nanometer-scaled coating that is deposited on a surface of said filaments, wherein said coating comprises a cathode active material capable of absorbing and desorbing lithium ions and said coating has a thickness less than 10 μm.*

Example 4: *A method for automatically surfacing tagged content adjunct to a vertical application, the method comprising: receiving and parsing text from content in an end user application; comparing the parsed text to social bookmarks and associated metadata from a social bookmarking system and matching portions of the content to respective ones of the social bookmarks and associated metadata based upon the comparison; and, directing a visual emphasis of the matched portions of the content in the end user application, whereby the end user application is unmodified to perform the receiving, comparing and directing steps.*

Example 5: *A storable foamable emulsion composition adapted for delivery of an active pharmaceutical ingredient (API) to a delivery site in a subject, the composition comprising: a) at least one organic carrier selected from the group consisting of a hydrophobic organic carrier, an organic polar solvent, an emollient and mixtures thereof, at a concentration of about 2% to about 50% by weight; b) at least one surface-active agent at a concentration of about 0.01% to about 5% by weight; c) at least one polymeric agent selected from the group consisting of a bioadhesive agent, a gelling agent, a film forming agent and a phase change agent, each in a concentration of about 0.01% to about 5% by weight; d) water; e) an effective amount of at least one API selected from the group consisting of a steroid, a steroid derivative, and combinations thereof; f) optionally, a further active agent; and g) a propellant at a concentration of about 3% to about 25% by weight of the total foamable composition, wherein, at ambient temperature, the storable foamable emulsion composition is shakable, is resistant to centrifugation at about 3000 rpm for about 10 min, is substantially devoid of crystals, is resistant to at least one freeze-thaw cycle and does not phase separate within at least about one month; wherein the at least one API remains chemically stable for at least about one month; and wherein the composition is stored in an aerosol container and upon release expands to form a breakable foam*

having an average bubble size range of about 30 to about 250 micron.

B Essential Legal Codes for Patent Examination

The followings are referred to in the paper that provision *patentability*:

B.1 35 U.S.C. 101 Inventions patentable.

Whoever invents or discovers any new and useful process, machine, manufacture, or composition of matter, or any new and useful improvement thereof, may obtain a patent therefor, subject to the conditions and requirements of this title.

B.2 35 U.S.C. 102 (a) Conditions for patentability; novelty.

(a) NOVELTY; PRIOR ART.—A person shall be entitled to a patent unless— (1) the claimed invention was patented, described in a printed publication, or in public use, on sale, or otherwise available to the public before the effective filing date of the claimed invention; or (2) the claimed invention was described in a patent issued under section 151 , or in an application for patent published or deemed published under section 122(b) , in which the patent or application, as the case may be, names another inventor and was effectively filed before the effective filing date of the claimed invention.

B.3 35 U.S.C. 103 Conditions for patentability; non-obvious subject matter.

A patent for a claimed invention may not be obtained, notwithstanding that the claimed invention is not identically disclosed as set forth in section 102 , if the differences between the claimed invention and the prior art are such that the claimed invention as a whole would have been obvious before the effective filing date of the claimed invention to a person having ordinary skill in the art to which the claimed invention pertains. Patentability shall not be negated by the manner in which the invention was made.

B.4 35 U.S.C. 112 (a) (b) Specification.

(a) IN GENERAL.—The specification shall contain a written description of the invention, and of the manner and process of making and using it, in such full, clear, concise, and exact terms as to enable any person skilled in the art to which it pertains, or with which it is most nearly connected, to make and use the same, and shall set forth the best mode contemplated by the inventor or joint inventor of carrying out the invention. (b) CONCLUSION.—The specification shall conclude with one or more claims particularly pointing out and distinctly claiming the subject matter which the inventor or a joint inventor regards as the invention.

C Example Approval Rates across Common Patent Classes

Table 6 demonstrates the variations of approval rates in different patent classess, ranging from 63.1% to 93.2%, indicating the inclusion of patent class feature is critical.

D 37 CFR 1.215 Patent Application Publication

(a) The publication of an application under 35 U.S.C. 122(b) shall include a patent application publication. The date of publication shall be indicated on the patent application publication. The patent application publication will be based upon the specification and drawings deposited on the filing date of the application, as well as the application data sheet and/or the inventor's oath or declaration. The patent application publication may also be based upon amendments to the specification (other than the abstract or the claims) that are reflected in a substitute specification under § 1.125(b), amendments to the abstract under § 1.121(b), amendments to the claims that are reflected in a complete claim listing under § 1.121(c), and amendments to the drawings under § 1.121(d), provided that such substitute specification or amendment is submitted in sufficient time to be entered into the Office file wrapper of the application before technical preparations for publication of the application have begun. Technical preparations for publication of an application generally begin four months prior to the projected date of publication. The patent application publication of an application that has entered the national stage under 35 U.S.C. 371 may also include amendments made during the international stage. See paragraph (c) of this section for publication of an application based upon a copy of the application submitted via the Office electronic filing system. (b) The patent application publication will include the name of the assignee, person to whom the inventor is under an obligation to assign the invention, or person who otherwise shows sufficient

Table 6: Example Approval Rates across Common Classes.

USPC Code	Application Counts	Approval Rate	Description
716	4425	63.10%	COMPUTER-AIDED DESIGN AND ANALYSIS OF CIRCUITS AND SEMICONDUCTOR MASKS
362	28054	75.59%	ILLUMINATION
257	151435	80.43%	ACTIVE SOLID-STATE DEVICES (E.G.,TRANSISTORS, SOLID-STATE DIODES)
375	44245	89.10%	PULSE OR DIGITAL COMMUNICATIONS
718	6848	93.17%	ELECTRICAL COMPUTERS AND DIGITAL PROCESSING SYSTEMS: VIRTUAL MACHINE TASK OR PROCESS MANAGEMENT OR TASK MANAGEMENT/CONTROL

proprietary interest in the matter if that information is provided in the application data sheet in an application filed under § 1.46. Assignee information may be included on the patent application publication in other applications if the assignee information is provided in an application data sheet submitted in sufficient time to be entered into the Office file wrapper of the application before technical preparations for publication of the application have begun. Providing assignee information in the application data sheet does not substitute for compliance with any requirement of part 3 of this chapter to have an assignment recorded by the Office. (c) At applicant's option, the patent application publication will be based upon the copy of the application (specification, drawings, and the application data sheet and/or the inventor's oath or declaration) as amended, provided that applicant supplies such a copy in compliance with the Office electronic filing system requirements within one month of the mailing date of the first Office communication that includes a confirmation number for the application, or fourteen months of the earliest filing date for which a benefit is sought under title 35, United States Code, whichever is later. (d) If the copy of the application submitted pursuant to paragraph (c) of this section does not comply with the Office electronic filing system requirements, the Office will publish the application as provided in paragraph (a) of this section. If, however, the Office has not started the publication process, the Office may use an untimely filed copy of the application supplied by the applicant under paragraph (c) of this section in creating the patent application publication.

E Example Rejections Data

The following excerpt entries (entry 0 to entry 6) are an example for office actions one application receives. These entries are part of the processed data ingested into a document-based DBMS. Office actions are where allowance and rejection decisions are formally made and sent to the applicant. The key `"submissionDate"` indicate the date when the office action is made. In the following 7 entries, 3 office action are involved, dated at 2017-11-2 (entries 0, 1, 4) 2018-06-28 (entries 3, 5), and 2018-12-27 (entries 2, 6.) Keys `"hasRej101"`, `"hasRej102"`, `"hasRej103"`, `"hasRej112"` indicate which are the legal sections raised and involved in the office action. Key `"legalSectionCode"` indicates which part of the rejections are covered in the office action with this entry. For example, for the office action made on 2017-11-29, legal sections *35 U.S. Code 102, 103, 112* are involved which shall spawn 3 entries. Entry 0 covers legal section code *112*. Entry 1 covers legal section code *103*. Entry 4 covers legal section code *102*. Then entry 4 describes that claim numbers 30,64,66,67, as indicated by key `"claimNumberArrayDocument"` are rejected due to *35 U.S. Code 102* on 2017-11-29. We utilize the merged data and inferred date to extract classification labels.

Entry 0:

```
{
    "obsoleteDocumentIdentifier" : "JAIGBW6RRXEAPX1",
    "groupArtUnitNumber" : "2174",
    "legacyDocumentCodeIdentifier" : "CTFR",
    "submissionDate" : ISODate("2017-11-29T00:00:00.000Z"),
    "nationalClass" : "715",
    "nationalSubclass" : "794000",
    "headerMissing" : NumberDecimal("0.0"),
    "formParagraphMissing" : NumberDecimal("0.0"),
    "rejectFormMissmatch" : NumberDecimal("0.0"),
    "closingMissing" : NumberDecimal("0.0"),
    "hasRej101" : NumberDecimal("0.0"),
    "hasRejDP" : NumberDecimal("0.0"),
    "hasRej102" : NumberDecimal("1.0"),
    "hasRej103" : NumberDecimal("1.0"),
    "hasRej112" : NumberDecimal("1.0"),
    "hasObjection" : NumberDecimal("0.0"),
    "cite102GT1" : NumberDecimal("0.0"),
    "cite103GT3" : NumberDecimal("0.0"),
    "cite103EQ1" : NumberDecimal("1.0"),
    "cite103Max" : NumberDecimal("3.0"),
    "signatureType" : NumberDecimal("3.0"),
    "actionTypeCategory" : "rejected",
    "legalSectionCode" : "112",
    "paragraphNumber" : "b",
    "claimNumberArrayDocument" : "67",
    "createUserIdentifier" : "ETL_SYS",
    "createDateTime" : "2019-10-19T22:12:26Z"
},
```

Entry 1:

```
{
    "obsoleteDocumentIdentifier" : "JAIGBW6RRXEAPX1",
    "groupArtUnitNumber" : "2174",
    "legacyDocumentCodeIdentifier" : "CTFR",
    "submissionDate" : ISODate("2017-11-29T00:00:00.000Z"),
    "nationalClass" : "715",
    "nationalSubclass" : "794000",
    "headerMissing" : NumberDecimal("0.0"),
    "formParagraphMissing" : NumberDecimal("0.0"),
    "rejectFormMissmatch" : NumberDecimal("0.0"),
    "closingMissing" : NumberDecimal("0.0"),
    "hasRej101" : NumberDecimal("0.0"),
    "hasRejDP" : NumberDecimal("0.0"),
    "hasRej102" : NumberDecimal("1.0"),
    "hasRej103" : NumberDecimal("1.0"),
    "hasRej112" : NumberDecimal("1.0"),
    "hasObjection" : NumberDecimal("0.0"),
    "cite102GT1" : NumberDecimal("0.0"),
    "cite103GT3" : NumberDecimal("0.0"),
    "cite103EQ1" : NumberDecimal("1.0"),
    "cite103Max" : NumberDecimal("3.0"),
    "signatureType" : NumberDecimal("3.0"),
    "actionTypeCategory" : "rejected",
    "legalSectionCode" : "103",
    "claimNumberArrayDocument" : "27,60,61,62,63,65",
    "createUserIdentifier" : "ETL_SYS",
    "createDateTime" : "2019-10-19T22:12:26Z"
},
```

Entry 2:

```
{
    "obsoleteDocumentIdentifier" : "JPY30LXURXEAPX0",
    "groupArtUnitNumber" : "2174",
    "legacyDocumentCodeIdentifier" : "CTFR",
    "submissionDate" : ISODate("2018-12-27T00:00:00.000Z"),
    "nationalClass" : "715",
    "nationalSubclass" : "794000",
    "headerMissing" : NumberDecimal("0.0"),
    "formParagraphMissing" : NumberDecimal("0.0"),
    "rejectFormMissmatch" : NumberDecimal("0.0"),
    "closingMissing" : NumberDecimal("0.0"),
    "hasRej101" : NumberDecimal("0.0"),
    "hasRejDP" : NumberDecimal("0.0"),
    "hasRej102" : NumberDecimal("0.0"),
    "hasRej103" : NumberDecimal("1.0"),
    "hasRej112" : NumberDecimal("0.0"),
    "hasObjection" : NumberDecimal("0.0"),
    "cite102GT1" : NumberDecimal("0.0"),
    "cite103GT3" : NumberDecimal("1.0"),
    "cite103EQ1" : NumberDecimal("0.0"),
    "cite103Max" : NumberDecimal("4.0"),
    "signatureType" : NumberDecimal("3.0"),
    "actionTypeCategory" : "rejected",
    "legalSectionCode" : "103",
    "claimNumberArrayDocument" : "27,60,61,65,68,69,70,73",
    "createUserIdentifier" : "ETL_SYS",
    "createDateTime" : "2019-06-02T07:22:43Z"
},
```

Entry 3:

```
{
    "obsoleteDocumentIdentifier" : "JIVT2WZ9RXEAPX4",
    "groupArtUnitNumber" : "2174",
    "legacyDocumentCodeIdentifier" : "CTNF",
    "submissionDate" : ISODate("2018-06-28T00:00:00.000Z"),
    "nationalClass" : "715",
    "nationalSubclass" : "794000",
    "headerMissing" : NumberDecimal("0.0"),
    "formParagraphMissing" : NumberDecimal("0.0"),
    "rejectFormMissmatch" : NumberDecimal("0.0"),
    "closingMissing" : NumberDecimal("0.0"),
    "hasRej101" : NumberDecimal("0.0"),
    "hasRejDP" : NumberDecimal("0.0"),
    "hasRej102" : NumberDecimal("0.0"),
    "hasRej103" : NumberDecimal("1.0"),
    "hasRej112" : NumberDecimal("0.0"),
    "hasObjection" : NumberDecimal("0.0"),
    "cite102GT1" : NumberDecimal("0.0"),
    "cite103GT3" : NumberDecimal("0.0"),
    "cite103EQ1" : NumberDecimal("0.0"),
    "cite103Max" : NumberDecimal("3.0"),
    "signatureType" : NumberDecimal("3.0"),
    "actionTypeCategory" : "rejected",
    "legalSectionCode" : "103",
    "paragraphNumber" : "a",
    "claimNumberArrayDocument" : "30,64,66",
    "createUserIdentifier" : "ETL_SYS",
    "createDateTime" : "2019-05-24T01:19:15Z"
},
```

Entry 4:

```
{
    "obsoleteDocumentIdentifier" : "JAIGBW6RRXEAPX1",
    "groupArtUnitNumber" : "2174",
    "legacyDocumentCodeIdentifier" : "CTFR",
    "submissionDate" : ISODate("2017-11-29T00:00:00.000Z"),
    "nationalClass" : "715",
    "nationalSubclass" : "794000",
    "headerMissing" : NumberDecimal("0.0"),
    "formParagraphMissing" : NumberDecimal("0.0"),
    "rejectFormMissmatch" : NumberDecimal("0.0"),
    "closingMissing" : NumberDecimal("0.0"),
    "hasRej101" : NumberDecimal("0.0"),
    "hasRejDP" : NumberDecimal("0.0"),
    "hasRej102" : NumberDecimal("1.0"),
    "hasRej103" : NumberDecimal("1.0"),
    "hasRej112" : NumberDecimal("1.0"),
    "hasObjection" : NumberDecimal("0.0"),
    "cite102GT1" : NumberDecimal("0.0"),
    "cite103GT3" : NumberDecimal("0.0"),
    "cite103EQ1" : NumberDecimal("1.0"),
    "cite103Max" : NumberDecimal("3.0"),
    "signatureType" : NumberDecimal("3.0"),
    "actionTypeCategory" : "rejected",
    "legalSectionCode" : "102",
    "paragraphNumber" : "b",
    "claimNumberArrayDocument" : "30,64,66,67",
    "createUserIdentifier" : "ETL_SYS",
    "createDateTime" : "2019-10-19T22:12:26Z"
},
```

Entry 5:

```
{
    "obsoleteDocumentIdentifier" : "JIVT2WZ9RXEAPX4",
    "groupArtUnitNumber" : "2174",
    "legacyDocumentCodeIdentifier" : "CTNF",
    "submissionDate" : ISODate("2018-06-28T00:00:00.000Z"),
    "nationalClass" : "715",
    "nationalSubclass" : "794000",
    "headerMissing" : NumberDecimal("0.0"),
    "formParagraphMissing" : NumberDecimal("0.0"),
    "rejectFormMissmatch" : NumberDecimal("0.0"),
    "closingMissing" : NumberDecimal("0.0"),
    "hasRej101" : NumberDecimal("0.0"),
    "hasRejDP" : NumberDecimal("0.0"),
    "hasRej102" : NumberDecimal("0.0"),
    "hasRej103" : NumberDecimal("1.0"),
    "hasRej112" : NumberDecimal("0.0"),
    "hasObjection" : NumberDecimal("0.0"),
    "cite102GT1" : NumberDecimal("0.0"),
    "cite103GT3" : NumberDecimal("0.0"),
    "cite103EQ1" : NumberDecimal("0.0"),
    "cite103Max" : NumberDecimal("3.0"),
    "signatureType" : NumberDecimal("3.0"),
    "actionTypeCategory" : "cancelled",
    "claimNumberArrayDocument" : "27,30,68,69,70,73",
    "createUserIdentifier" : "ETL_SYS",
    "createDateTime" : "2019-05-24T01:19:15Z"
},
```

Entry 6:

```
{
    "obsoleteDocumentIdentifier" : "JPY30LXURXEAPX0",
    "groupArtUnitNumber" : "2174",
    "legacyDocumentCodeIdentifier" : "CTFR",
    "submissionDate" : ISODate("2018-12-27T00:00:00.000Z"),
    "nationalClass" : "715",
    "nationalSubclass" : "794000",
    "headerMissing" : NumberDecimal("0.0"),
    "formParagraphMissing" : NumberDecimal("0.0"),
    "rejectFormMissmatch" : NumberDecimal("0.0"),
    "closingMissing" : NumberDecimal("0.0"),
    "hasRej101" : NumberDecimal("0.0"),
    "hasRejDP" : NumberDecimal("0.0"),
    "hasRej102" : NumberDecimal("0.0"),
    "hasRej103" : NumberDecimal("1.0"),
    "hasRej112" : NumberDecimal("0.0"),
    "hasObjection" : NumberDecimal("0.0"),
    "cite102GT1" : NumberDecimal("0.0"),
    "cite103GT3" : NumberDecimal("1.0"),
    "cite103EQ1" : NumberDecimal("0.0"),
    "cite103Max" : NumberDecimal("4.0"),
    "signatureType" : NumberDecimal("3.0"),
    "actionTypeCategory" : "cancelled",
    "claimNumberArrayDocument" : "27,30,68,69,70,71,72,73",
    "createUserIdentifier" : "ETL_SYS",
    "createDateTime" : "2019-06-02T07:22:43Z"
}
```

F Full ROC Figures

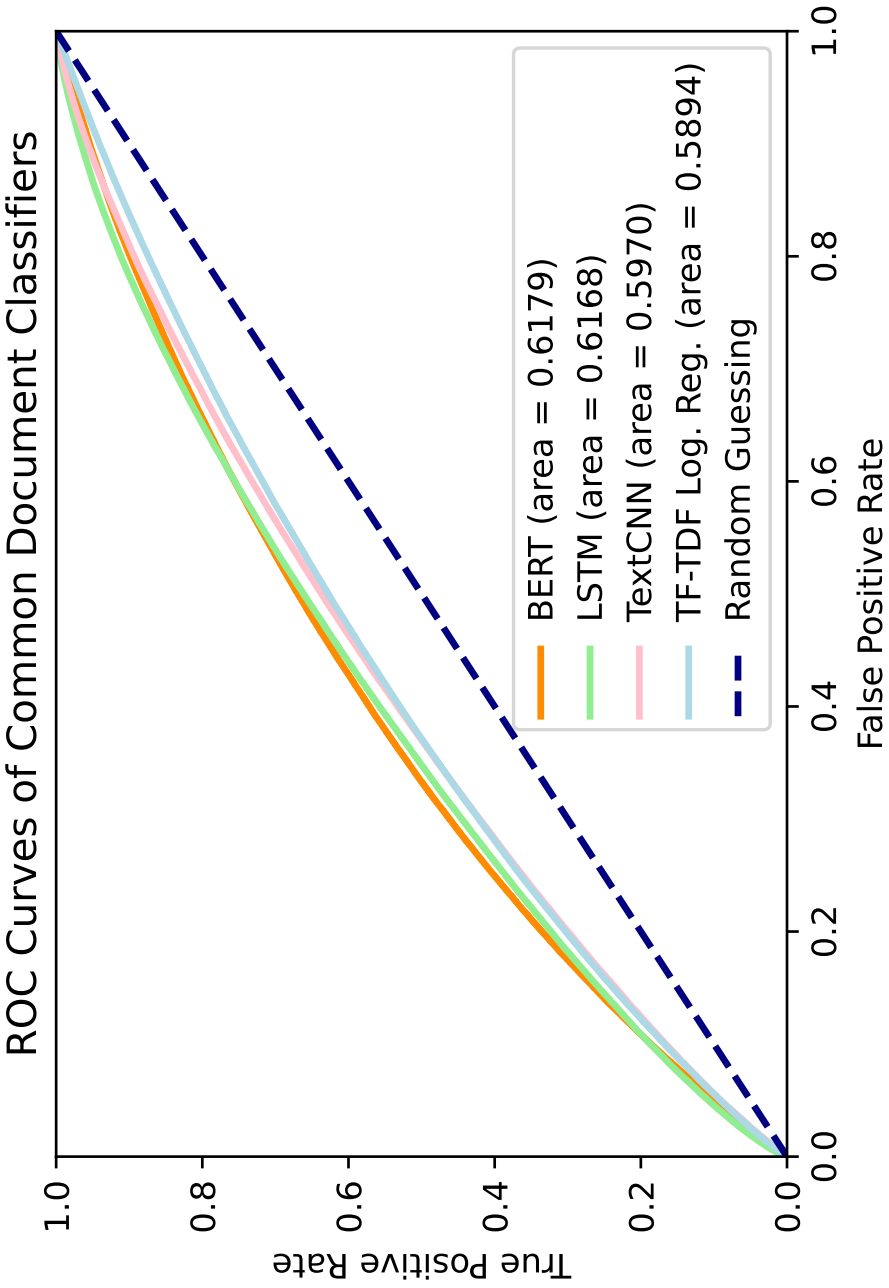


Figure 4: ROC Curves of Common Document Classifiers. BERT and LSTM are arguably the most effective.

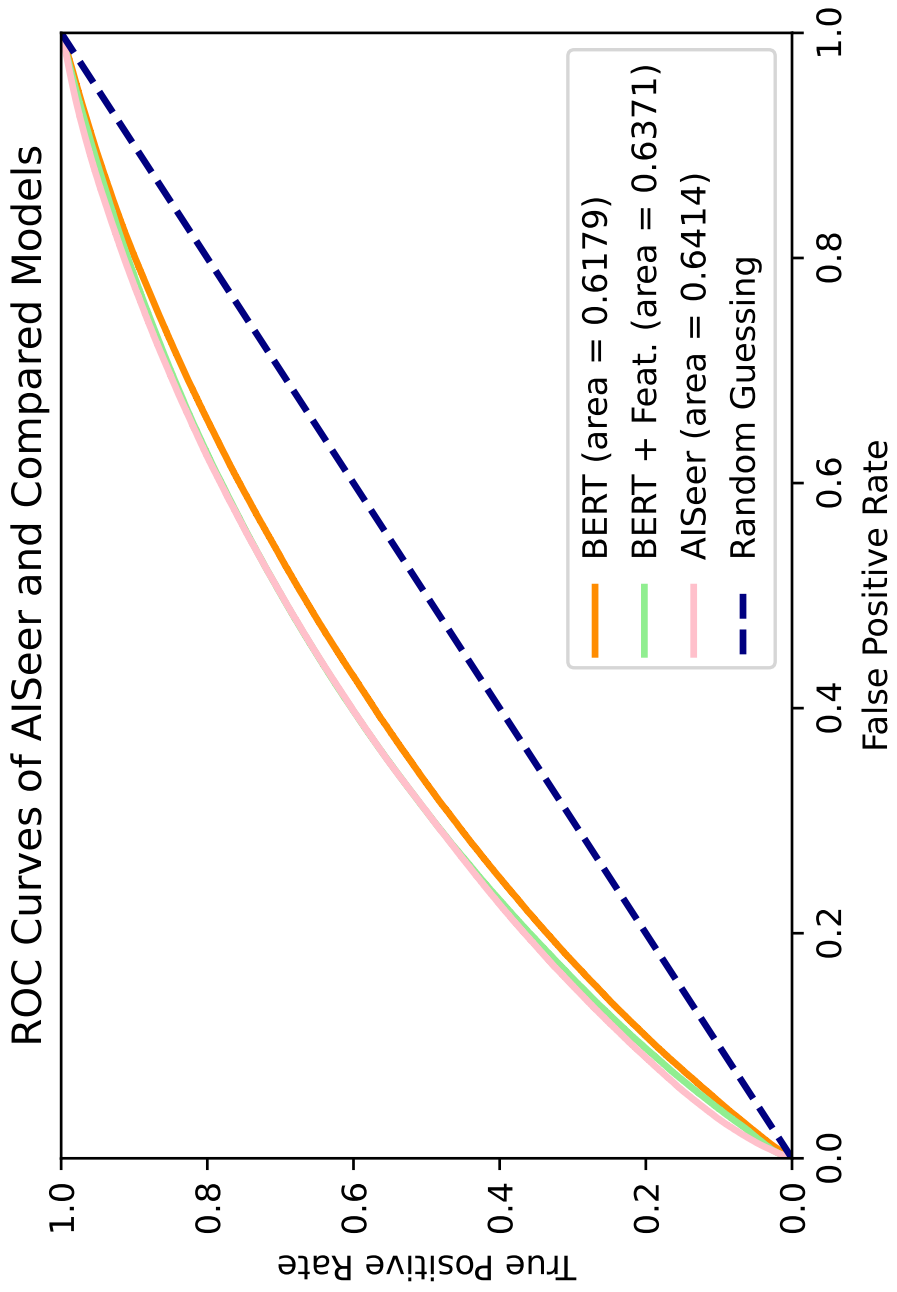


Figure 5: ROC Curves for AISeer and Compared Models.