Volume 5, No. 5 | September–October 2022

RAIL

The Journal of Robotics, Artificial Intelligence & Law

fastcase FULL COURT PRESS

RAIL
The Journal of Robotics, Artificial Intelligence & Law
Volume 5, No. 5 | September–October 2022

THE JOURNAL OF ROBOTICS, ARTIFICIAL INTELLIGENCE & LAW (ISSN 2575-5633 (print) /ISSN 2575-5617 (online) at $495.00 annually is published six times per year by Full Court Press, a Fastcase, Inc., imprint. Copyright 2022 Fastcase, Inc. No part of this journal may be reproduced in any form—by microfilm, xerography, or otherwise—or incorporated into any information retrieval system without the written permission of the copyright owner. For customer support, please contact Fastcase, Inc., 711 D St. NW, Suite 200, Washington, D.C. 20004, 202.999.4777 (phone), 202.521.3462 (fax), or email customer service at support@fastcase.com.

Publishing Staff
Publisher: Morgan Morrissette Wright
Production Editor: Sharon D. Ray
Cover Art Design: Juan Bustamante

Cite this publication as:

The Journal of Robotics, Artificial Intelligence & Law (Fastcase)

A Full Court Press, Fastcase, Inc., Publication

Editorial Office

711 D St. NW, Suite 200, Washington, D.C. 20004
https://www.fastcase.com/

POSTMASTER: Send address changes to THE JOURNAL OF ROBOTICS, ARTIFICIAL INTELLIGENCE & LAW, 711 D St. NW, Suite 200, Washington, D.C. 20004.

Articles and Submissions

Direct editorial inquiries and send material for publication to:

Steven A. Meyerowitz, Editor-in-Chief, Meyerowitz Communications Inc., 26910 Grand Central Parkway, #18R, Floral Park, NY 11005, smeyerowitz@meyerowitzcommunications.com, 631.291.5541.

Material for publication is welcomed—articles, decisions, or other items of interest to attorneys and law firms, in-house counsel, corporate compliance officers, government agencies and their counsel, senior business executives, scientists, engineers, and anyone interested in the law governing artificial intelligence and robotics. This publication is designed to be accurate and authoritative, but neither the publisher nor the authors are rendering legal, accounting, or other professional services in this publication. If legal or other expert advice is desired, retain the services of an appropriate professional. The articles and columns reflect only the present considerations and views of the authors and do not necessarily reflect those of the firms or organizations with which they are affiliated, any of the former or present clients of the authors or their firms or organizations, or the editors or publisher.

QUESTIONS ABOUT THIS PUBLICATION?

For questions about the Editorial Content appearing in these volumes or reprint permission, please contact:

Morgan Morrissette Wright, Publisher, Full Court Press at mwright@fastcase.com or at 202.999.4878

For questions or Sales and Customer Service:

Customer Service
Available 8 a.m.–8 p.m. Eastern Time
866.773.2782 (phone)
support@fastcase.com (email)

Sales
202.999.4777 (phone)
sales@fastcase.com (email)
ISSN 2575-5633 (print)
ISSN 2575-5617 (online)

A Deep Learning Model for Predicting Patent Applications Outcomes

Oscar A. Garcia, Naisargi Dave, Qie Tang, Josvin John, Anthony Topper, Kashyap Bhuva, Manasi Shrotri, Sayali Shelke, Xiaosong Wen, Dr. Reza Mollaaghababa, Prof. Fatemeh Emdad, Prof. Chun-Kit Ngan, Prof. Elke Rundensteiner, and Prof. Seyed A Zekavat*

Natural language processing has evolved over time and is now used widely for text classification. This article focuses on building a natural language processing classification model to determine whether the U.S. Patent and Trademark Office ("USPTO") will affirm or reverse an appealed patent application rejection. Appeals from USPTO rejections of patent applications are a critical component of patent prosecution practice. The model is trained on affirmed and reversed claims before 2018 that are available on the USPTO and Patent Application Information Retrieval websites.

A patent is an intellectual property right that is granted by the U.S. Constitution to inventors. A patent allows an inventor to exclude others from making, selling, or using an invention for a limited duration. The process of obtaining a patent involves filing a patent application with the U.S. Patent and Trademark Office ("USPTO"). The application undergoes examination at the USPTO and if it satisfies various statutory requirements for patentability, it matures into a patent.

The claims of a patent define the scope of protection provided by the patent. Typically, the initially filed claims undergo changes during the examination process before the issuance of a patent. Such changes generally arise through negotiations between an applicant (and in most cases an applicant's patent attorney) and a patent examiner assigned to examine the patent application.

An applicant also has the opportunity to appeal an examiner's final rejection to the Patent Trial and Appeal Board ("PTAB"). The PTAB can affirm, reverse, or remand an examiner's decision with respect to a rejected claim or send the case back to the examiner for further consideration. We also note that under the America

Robotics, Artificial Intelligence & Law / September–October 2022, Vol. 5, No. 5, pp. 347–355.

ISSN 2575-5633 (print) /ISSN 2575-5617 (online)

Invents Act ("AIA"), which came into effect in 2012, a third party can challenge the validity of an issued patent before the PTAB.

In this article, we report the results of the application of a "BERT" algorithm for predicting whether the PTAB will affirm or reverse an examiner's decision. We employed a number of application programming interfaces ("APIs") that the USPTO has made available to access the information we needed for input into the machine learning algorithms. In particular, we used an API known as BulkAPI to access data corresponding to claim texts and the disposition of the respective patent applications.

We used unique identifiers associated with a filed patent application for extracting the required information. The information provided by the APIs was in the form of json and xml files. We aggregated the data by connecting data points using common identifiers, such as, patent application numbers.

Related Work

This section of the article briefly reviews the recent natural language processing contributions for application to patent data.

David Winer[1] developed algorithms to predict two major outcomes: whether the PTAB will institute a trial based on a post-grant challenge brought against one or more claims of a patent, and whether the challenged claims will be invalidated if a trial is instituted. This article focuses on the former algorithm, that is, whether the PTAB would institute a trial (we refer to the PTAB's decision in this regard as acceptance/denial where acceptance refers to the PTAB's decision to institute a trial and denial refers to the PTAB's decision not to institute a trial).

To predict acceptance/denial, we used natural language processing ("NLP") such as Word2Vec to convert each litigated patent document into thousands of numeric features. Following the combination of these text-based features with patent metadata, we used two primary machine learning algorithms to attempt to classify these documents as belonging to acceptance or denial categories: support vector classification and random forests. Our best performing model was the simple linear support vector classifier with $1/\lambda$ set to 10, giving an accuracy of 78 percent on test data, precision value 0.82, and recall value 0.87.

PatentBERT[2] focused on fine-tuning a pre-trained BERT model and applying it to patent classification; that is, organizing patents into specific categories according to their technical content. When applied to large data sets of over two million patents, this approach outperformed the state-of-the-art approach using the Convolutional Neural Network ("CNN") with word embeddings. In addition, this approach focuses on patent claims without considering other portion of the patent documents. DeepPatent, also a deep learning algorithm for patent classification based on CNN and word vector embedding, was proposed by Li Shaobo and others[3] and represents an improvement over PatentBERT.

Background

Bidirectional Encoder Representations from Transformers ("BERT")[4] is a language model published by Google. This is a bidirectionally trained model, which has a deeper sense of context and flow than a single direction language model. BERT learns information from left to right and right to left. Such bidirectional learning makes BERT more powerful in making accurate predictions than left-to-right models. It uses Masked LM ("MLM") to mask words in a sentence and then it tries to predict them. This enables the model to use the full context of the sentence and take both previous and next tokens into account at the same time. Use of these pre-trained word embeddings from BERT will help improve machine understanding of claims.

Data Set and Preprocessing

Data Collection

All of the relevant data is provided publicly through the USPTO. As noted above, the USPTO provides a variety of APIs at the domain developer.uspto.gov, including APIs for accessing the textual content of a patent application, along with its application status. The "Bulk" API provides access to archived records of all patent applications, packaged into zip files organized by week. The PTAB Proceedings and Documents REST API can provide the decision status for all appealed patents. Our approach required obtaining relevant data from these distinct sources and then matching records appropriately. There are several kinds of identifiers used by the USPTO for labeling patents, but some of them are not unique and some are not provided by certain APIs. We found that the patent application number (named as "respondentApplicationNumberText" in the API) was the most reliable to match entries. The data consisted of patent application documents and their corresponding PTAB decisions from 2002 to 2016.

Data Preprocessing

Transformer-based models, like BERT, are advanced enough to gain information from stop words and lemmas, which are commonly removed in traditional NLP models. In contrast, the preprocessing in BERT involves only removing special characters that are introduced from parsing the application document from xml files.

Use Case

After a claim of a patent application has been rejected twice or has been finally rejected by an examiner, the applicant may seek to appeal the decision to the PTAB, which will reverse or affirm the examiner's decision or remand the decision back to the examiner for further consideration. We present models that attempt to predict whether the PTAB will reverse or affirm the examiner's decision.

Because the reversal of the examiner's decision (the "Reversed class") by the PTAB represents a positive outcome for the applicant, we have concentrated on optimizing the prediction accuracy for the Reverse class at the expense, in some cases, of sacrificing some correct predictions of the affirmed class. For this particular use case, we propose adding a bias to the BERT model to increase the correct classifications of the Reversed class.

Methodology

In addition to the text of the claims in a patent application, we also added additional information as features. During the examination process, the examiner may reject a claim based on multiple grounds. During appeal, the PTAB may agree with some of the grounds of rejection raised by the examiner while disagreeing with others. However, it is sufficient for the PTAB to agree with only one rejection ground for the claim not to be allowed.

Although the USPTO API can be used to extract the reasons an examiner rejected a claim of a patent application that is under appeal, the API cannot be employed to extract the PTAB's decision with respect to each ground of rejection raised relative to a claim. For example, the PTAB may not agree with an examiner that a claim is anticipated by a cited reference but may agree with the examiner that the claim is obvious with respect to that reference alone or in combination with other references. Therefore, we did not attempt to build a model that can predict the opinion of the PTAB for each ground of rejection, but only whether the PTAB's final decision led to affirmance of the totality of the examiner's rejection of a claim or its reversal.

We used the Hugging face library,[5] which contains many pertinent NLP models and an easy to use interface for building our model. More specifically, we used the "bert-base-cased" model and then fine-tuned it to our specific task. The issues raised by the examiner and the applications assigned art unit were added as one-hot encoded features concatenated to the final layer of the BERT embedding to form the input to the classification head.

The classification head had two hidden layers with 512 and 256 neurons, respectively, with ReLU activation functions; only these neurons were trained using the AdamW optimizer with a learning rate of 1e-5 and Binary Cross Entropy loss, while the BERT embeddings were frozen to avoid modifying their inherent language understanding. The output probabilities of the models were calculated using a sigmoid function; a constant bias was added here to shift the probabilities toward the Reversed class (the result of this bias is reported separately from the unbiased model).

This model was trained using an early stopping strategy with a patience of three and a delta of 0.001. A stratified sample of 70 percent for training, 15 percent for testing that controls the early stopping, and 15 percent for validation was used. This experiment was performed five times with different random seeds to ensure the validity of the results; their average is presented.

Results

The average result among all unbiased BERT iterations can be seen in Figure 1. This confusion matrix corresponds with a 59.35

percent accuracy and a 0.5929 ROC[6] with a balanced prediction power between the two classes.

The results for the biased models were not consistent when using a single bias value; this means that bias values have to be calibrated for each iteration to reach the desired prediction levels. A single biased model, which included a bias of 0.035 toward the Reversed class, is presented as an example in Figure 2. This biased model had a lower accuracy of 0.56 percent but a higher ROC of 0.5939.

Figure 1. Unbiased BERT Confusion Matrix

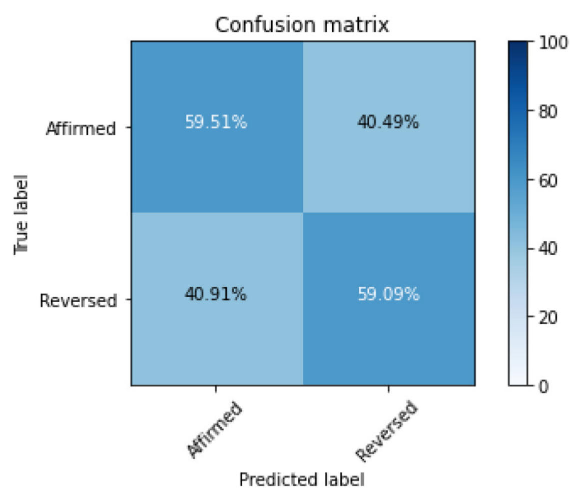
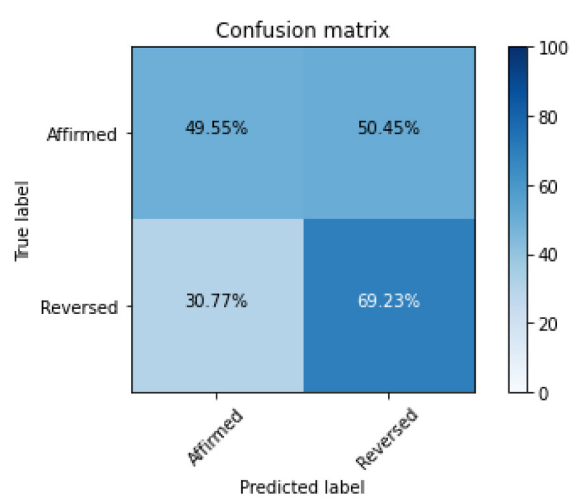


Figure 2. Biased BERT Confusion Matrix



To better understand the implications of our models for the use case presented above, we show our results in easy to understand visualization in Figures 3 and 4 using 100 cases as an example, where the numbers displayed are rounded to the nearest integer.

Figure 3. Results Visualizations of the Unbiased BERT Model

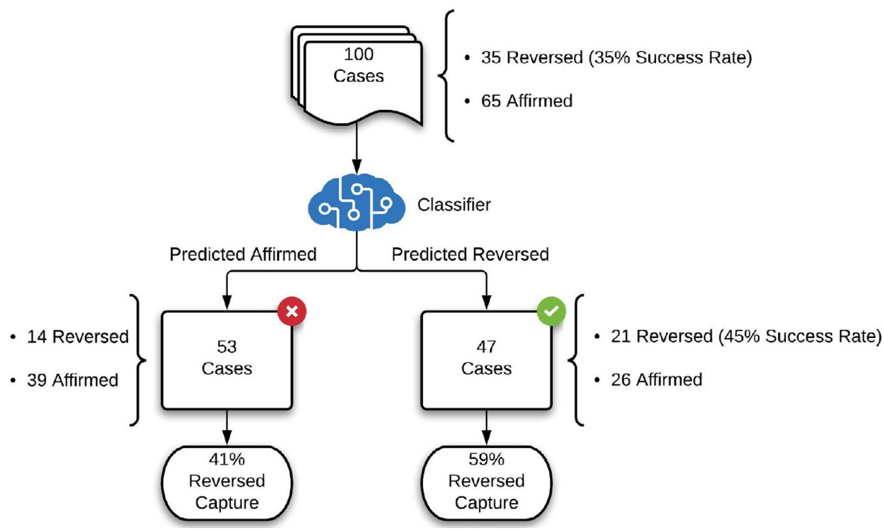
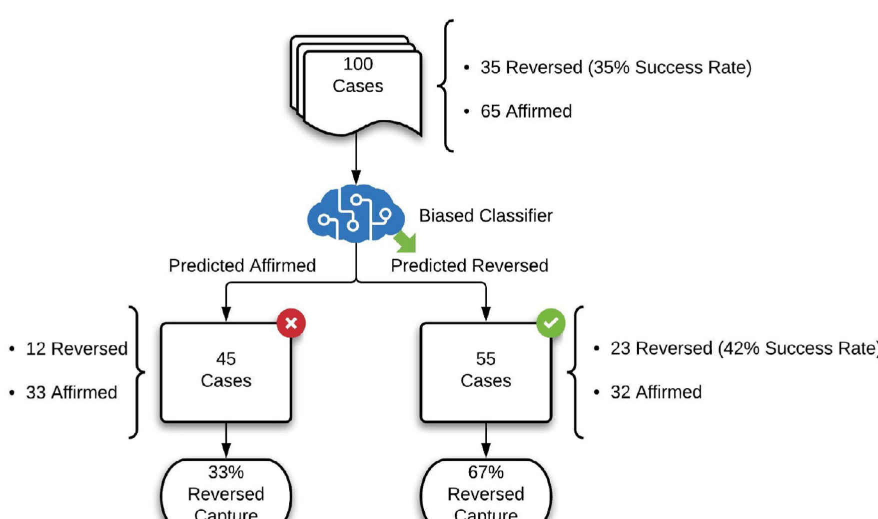


Figure 4. Results Visualizations of the Biased BERT Model



Conclusion

The difficulty of achieving a high performance could be because patent applications are highly technical texts that can be unintelligible to anyone who lacks the correct technical background even if they are highly educated and have good reading comprehension. Another reason classification may be difficult could be due to the numerous and very different ways a patent may be rejected or an examiner's decision affirmed; if the PTAB API would provide the board decisions for each issue raised by the examiner as discussed above, the prediction of the outcome could be done piecemeal by more specialized models, leading to a better performance.

Our BERT model does consistently better than random chance and we believe that it can assist a human agent to identify promising patent applications that are likely to be reversed by the PTAB, especially when using our biased BERT model.

Notes

* All of the authors are currently or formerly associated with Worcester Polytechnic Institute except for Dr. Reza Mollaaghababa, who is a partner at Nelson Mullins Riley & Scarborough, where he focuses his practice on intellectual property law. The authors may be contacted at ogarciafernandez@wpi.edu, ndave@wpi.edu, qtang@wpi.edu, jtjohn@wpi.edu, ajtopper@wpi.edu, kbbhuva@wpi.edu, mvshrotri@wpi.edu, sbshelke@wpi.edu, xwen2@wpi.edu, reza.mollaaghababa@nelsonmullins.com, femdad@wpi.edu, cngan@wpi.edu, rundenst@wpi.edu, and rezaz@wpi.edu, respectively.

1. David Winer. Predicting bad patents: Employing machine learning to predict post-grant review outcomes for us patents. Master's thesis, EECS Department, University of California, Berkeley, May 2017.

2. Jieh-Sheng Lee and Jieh Hsiang. PatentBERT: Patent classification with fine-tuning a pre-trained BERT model. *World Patent Information*, 61(101965), 2020.

3. Shaobo Li, Hu Jie, Cui Yuxin, and Jianjun Hu. DeepPatent: Patent classification with convolutional neural networks and word embedding. *Scientometrics*, 117(721-744), 2018.

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

5. Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine

Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander M. Rush. Hugging-face's transformers: State-of-the-art natural language processing, 2020.

6. *I.e.*, "Receiver Operating Characteristic" curve.