THE C. BOYDEN GRAY
Center *for the* Study *of the* Administrative State
ANTONIN SCALIA LAW SCHOOL • GEORGE MASON UNIVERSITY

Using Deep and Active Learning Classifiers to Identify Congressional Delegation to Administrative Agencies

Joshua Y. Lerner
Gregory P. Spell

CSAS Working Paper 21-31

GEORGE MASON UNIVERSITY | ANTONIN SCALIA LAW SCHOOL

Using Deep and Active Learning Classifiers to Identify Congressional Delegation to Administrative Agencies*

Joshua Y. Lerner[1] and Gregory P. Spell[2]

[1]Data Scientist and Research Methodologist, NORC at the University of Chicago

[2]Department of Electrical and Computer Engineering, Duke University

July 6, 2021

Abstract

Congressional oversight of the federal bureaucracy remains key to understanding implementation of the law. Essential to this are theories of how and why Congress delegates powers to administrative agencies. Using an active learning convolutional neural network on bill text, we classify bill sections by their role in delegating to administrative agencies, applying an iteratively improving coding scheme that enhances existing supervised learning approaches. We systematically study the statutory scope of administrative agencies and develop a first-of-its-kind dataset to study how delegation develops. First, we benchmark our measure against existing proxies for delegation. We then find evidence that, as bills advance through the legislative process, the delegation scope winnows. We also find that traditional expectations about unified versus divided government matter less when looking at all legislation, confounding the well-known ally principal. We conclude with a discussion of delegatory scope and other extensions of our method to new data.

Word Count: 9875

*We would like to thank the following people for their helpful comments on earlier versions: Mathew McCubbins, John Aldrich, Kristen Renberg, Kiran Auerbach, Hannah Ridge, Robert Shaffer, Justin Grimmer, Michael Crespin, Sarah Bouchat, Austin Bussing, and attendees at PolMeth 2018 and APSA 2018.

Delegation of powers from political actors to agents or agencies tasked with enacting policies remains at the heart of the challenges of modern governance. Even beyond the formal substance of the powers, any delegation has two essential elements that a coherent measurement approach must deal with: the identity of the agents who hold policy-making authority due to the delegation and any action restricting or enabling the exercise of that authority. For example, Congress may designate that the Environmental Protection Agency shall promulgate rules and regulations to protect endangered species, but only after issuing the rules and allowing for public comment. As important as it is, delegating legislation is not the only kind of law enacted by governments, and researchers have applied selection criteria to yield a small sample on which to apply a labor-intensive coding framework.

While this conceptualization is well known, applications of this concept have always run into a problem of measurement. Delegation is fundamentally a textual act: the legislature restricts or enables an agency in their actions through written law and is codified subsequently in oversight hearings, mark-up procedures, court cases, and more. Consequently, even the most rudimentary measurement of delegation has to grapple with the written word, either directly or indirectly, to capture what in particular is occurring. In any individual piece of legislation, this task is straightforward; identifying the agency and the change to their authority requires only a close reading of the bill. However, this approach has severe limitations for large-scale applications and empirical studies, namely the time and effort such an undertaking would require.

Innovations in machine learning and natural language processing simplify scaling up these connections. In particular, these methods are capable of predicting qualitative labels from quantifiable features. After being trained on a set of hand-labeled examples, the learning algorithm is deployed to predict labels for a much larger data set, significantly reducing the up-front costs of setting up a human-run labeling project. However, most machine learning operations separate the hand-coding from the evaluation process and treat it merely as a single attempt at attaining accuracy. Complex concepts will be hard to translate into a simple coding scheme, and thus machine coding done this way will can have additional problems.

In this paper, we argue that researchers can effectively combine their expertise through an interactive machine-learning framework that will, in essence, learn on the go. More commonly referred to as an "active learning" approach to classification, we argue that this framework best combines the portability and flexibility of machine learning with the expert evaluation usually used in more modestly defined approaches, such as case studies.

To demonstrate the utility of this framework, we tackle a canonical problem in political science: how and when Congress delegates authority to administrative agencies. We create a new dataset on legislative delegation. We address Congressional delegation for three reasons. First, it is an important problem, as delegation facilitates the implementation of laws written by Congress and the bureaucracy represents the most direct way any average citizen interacts with the federal government.

Second, theories of delegation in Congress abound in political science, economics, and public administration, but there have been few large-n empirical tests of these theories. There are numerous competing, highly comprehensive theories about how, why, and when Congress delegates authority, but up until now, most studies either focused only on a single bit of policy (Huber & Shipan 2002), a handful of "significant" bills (Epstein & O'Halloran 1999), or forewent empirical analysis altogether and focused on model building (see Gailmard & Patty 2012, for a comprehensive overview of such models). Having a measure that is both broad and readily comparable should allow us to evaluate parts of these theories as well as further research into the relationship between Congress and the bureaucracy. Since we include bills as they advance through the legislative process, this provides the first analysis on how delegation changes throughout and allows for a reexamination of the effect of divided and unified government on delegation.

This paper proceeds as follows: first, we discuss the delegation of authority from Congress to administrative agencies and explain why it has always posed a unique measurement challenge. Second, we address this measurement problem by using a deep and active learning model, with discussions of what makes this different from other machine learning approaches. We then assess classification accuracy and discuss the performance of our active learning model, how it improves upon existing models, and what can be done with our newly labeled dataset. We conclude by analyzing the properties of bills that delegate authority, test theories of partisanship and lawmaking in Congress, and compare our direct measure of delegation with existing proxies commonly used in the literature.

Congressional Delegation to Administrative Agencies

Delegation is a necessity for the operation of any large-scale enterprise, least of all governments. Political scientists generally view congressional delegation as a trade-off between efficiency and accountability (Kiewiet & McCubbins 1991; Epstein & O'Halloran 1999). Hypothetically, congressional delegation can have enormous productivity gains for both individual members of Congress and administrative agencies. However, in the course of designing delegation, Congress faces numerous internal coordination problems. Further, Congress must account for how delegation can create opportunities for agencies to act against its interests — standard in all principal-agent models. For these reasons — the institutional, partisan, and policy-making nature — congressional delegation has been a focal point for the study of political institutions. Scholars have formalized the intuition behind delegation through versions of the "ally principle" (Epstein & O'Halloran 1999; Huber & Shipan 2002; Moe 2012; Farhang & Yaver 2016), which argues that when the executive's interests are aligned with those of the legislature, legislators are more willing to pass legislation that delegates significant authority to agencies. By contrast, when legislative and executive policy interests diverge, legislators favor institutional structures that provide increased oversight opportunities.

Beyond institutional factors, some have posited that the design of agency authority is affected by characteristics of the issues and policy areas addressed (McCubbins 1985; Epstein & O'Halloran 1999). While the role of Congressional process and policy areas has been discussed extensively in the theoretical literature on allocating authority to agencies, these concepts are understudied in empirically oriented scholarship. This limitation results from a measurement problem, even for motivated researchers: reading and interpreting legal texts is labor-intensive. Thus, most empirical work on the allocation of authority has been restricted to single policy areas or to small sets of "significant" legislation. Given the limitations of previous empirical studies on delegation, our work to create a generalized dataset measuring delegation should fill an essential hole in the literature and provide fertile ground for the testing of new and old theories.

Difficulty of Measuring Delegation

While delegation has always been an important topic in political science, how it has been measured has attracted some controversy. Through most of the 1980s, Congressional delegation to the executive branch mainly was evidenced through the use of case studies; close textual readings of bills, but nothing directly lending itself to a measurement strategy (McCubbins *et al.* 1987, 1989; Kiewiet & McCubbins 1991).

Epstein & O'Halloran (1996) pushed the literature to measure administrative discretion in individual acts of delegation by Congress by comparing the number of provisions that grant and constrain the President's authority. They differentiate fourteen categories of constraints often imposed upon agents, which include: limits on agent's power to expend resources, actions that require pre-approval by another actor, legislative veto power over regulatory changes, ex-ante consultation (including approval) or ex-post reporting requirements, and specified processes for rulemaking. However, they do not compare the importance of these constraints. Another problem with the Epstein & O'Halloran (1999) approach is that it did not directly scale: they were only able to look at a small subset of legislation that became law. For as comprehensive as their book is, they only analyze 262 bills.

The next major development of delegation as a measurement comes from Huber & Shipan (2002), who use bill length — the total number of words in a bill — as a proxy for how much detail the legislature leaves for the administrative agency. They assert that this measure captures how much discretion is left for agencies to interpret the bill and freedom in their action. Huber & Shipan (2002) claim that longer bills are more likely to require specific returns from agencies, and this generally entails more fine-grained oversight and control. Shorter bills, ones that discuss delegation and discretion more broadly, are less likely to have the same requirements and instead reflect less deference from the legislature. It has been accepted by the literature (see Clinton *et al.* (2012) as an example) that longer bills mean added delegation.

In some ways, our approach marries the generalizability of Epstein & O'Halloran (1999) and Huber & Shipan (2002) with the explicitly close reading based approaches of earlier studies, in particular McCubbins *et al.* (1987) and Kiewiet & McCubbins (1991). Because of the textual nature of the task, we expect the classification of delegation in legislation to work well in a deep and active learning environment. As will be discussed later, active learning is most often utilized in machine learning settings where getting additional training labels is costly. In this case, we can see precisely why hand-coding delegation in bills would be an apropos application of this approach. First, it is a task that requires both training and familiarity with how Congress writes bills: the use of statutory language is always done deliberately and with the mind that the courts, the Executive Branch, and future Congress will have to interpret specifically what was written (McNollgast 1994). Furthermore, it is in Congress's best interest to standardize the writing of bills such that little is open to interpretation (McCubbins *et al.* 1987).

Secondly, although relying on the uniformity of statutory language, this task still requires a fair amount of careful reading, given the agencies' idiosyncrasies. This uncertainty, combined with the simple classification scheme, gives us a situation where coding any given section as delegatory is labor-intensive but straightforward to scale up the coding. Thus, the need for active learning: if we achieve reasonable accuracy in this classification task without hand-coding too many sections identifying which sections improve accuracy efficiently is critical. We believe that this exact issue is not unique to coding agency delegation in Congress but would apply to many classification tasks measuring latent concepts. This approach reduces the need for additional computational bottlenecks for such tasks and could make classifying abstract features in large documents much more tractable.

An Active Learning Convolutional Neural Network for Classifying Text

It is impossible to think of our most important legal and political concepts without relying on the written language surrounding them. It is not surprising, then, that the study of political language has expanded tremendously over the last decade. In particular, research into text-as-data methods in the social sciences has grown exponentially, as discussed in Grimmer & Stewart (2013) and Wilkerson & Casas (2017).

Classification is one of the most popular objectives in all text-as-data work. Generally falling under the umbrella of supervised learning, classification tasks take a set of inputs, along with their corresponding outputs (also called labels), with the goal of discovering some underlying map between input and output. Supervised methods often require thousands of training examples, rendering them a non-starter for many researchers and projects. However, there are often creative ways to reduce the effort required. We see an example related to our own in Anastasopoulos & Bertelli (2020), who use existing classifications of agency delegation in EU legislation to perform supervised labeling of other years not covered in the dataset.

Active learning — also called "query learning" in computer science — is a machine learning approach to bolster classification performance by selecting (or "querying") further examples to use in model training. This querying allows an algorithm to obtain higher classification accuracies as training data is added than if new examples were instead randomly chosen. An active learner typically poses its queries over unlabeled data, which are then labeled by a human annotator and subsequently added to the dataset. Active learning is well-motivated in many modern machine learning problems, where unlabeled data may be abundant, but labels are complicated, time-consuming, or expensive to obtain (Settles 2009; Miller *et al.* 2019).

Researchers may find that they need to frequently retrain or update their models, particularly given new information they discover in the course of their exploration. Active learning offers a judicious way to update with new examples. We provide a motivating example in identifying delegation to administrative agencies. The general framework for delegating authority is relatively straightforward: some agency is given a task — often told it "shall" or "must" do something. A standard learner would identify the agency named and the verbs specifying the task and then use that information to identify most instances of delegation. However, there are hundreds of currently active agencies and programs, some of which have unusual names (the "Corporation for National and Community Service" as an example), that would reduce the likelihood of delegation being properly identified.[1]

We encountered this problem early on when discussions of the "Attorney General" were frequently mislabeled because, up until that point, there were no observations in our training data with a cabinet-level secretary referred to as anything but "Secretary." Because the specific words "Attorney General" have few analogous positions in other departments, it would have to be hand-coded explicitly for the model to learn what that is. Generally, this is not a complicated fix: hand label some of these aberrant observations, and it should solve the problem. However, what we could not know *a priori* was what exact issues were going to appear: moving to an interactive labeling and machine-learning framework alleviated these concerns because the classifier would be able to tell us where it was having difficulties discriminating between classes. Learning on the go was the preferred option because, though our classification scheme is simple, there are enough moving parts that it is hard to know

[1] Unless that agency also showed up in the training data which, given that there are hundreds of agencies, is likely to miss many.

precisely what problems would have arisen before we started the coding.[2]

In a typical supervised machine learning environment, a model is trained on a training partition of the data and evaluated on a testing partition. Given the performance, applied researchers may then decide to "deploy" the model for some practical purpose: for instance, the automated labeling of new examples. They may also determine that the model performance, as evaluated on the test data, was insufficient, and so they may choose to label more examples for model training. For our case, we argue that active learning should be used when selecting new examples to label, with the criterion for selection being the model's uncertainty in labeling new examples. After querying these uncertainly labeled documents from the model, they may be manually labeled by the researchers and appended to the training dataset. The process of training, evaluation, and querying then repeats until model evaluation suggests the model is robust enough to be deployed to label all remaining examples automatically.

As mentioned above, our criterion for actively querying new data examples is the model's uncertainty in label prediction — "uncertainty sampling." In the case of classification, this strategy amounts to identifying the observations closest to the classification boundary, such as examples with the smallest margin in max-margin models (e.g., SVM) or examples with logits closest to 0.5 for logistic regression. Other common querying strategies are ensemble approaches — wherein queries are made by identifying observations for which there is classification disagreement between multiple models — and expected-model-change approaches, in which examples are selected that would most significantly change the current model (Settles 2009). For our problem, we have implemented uncertainty sampling, as it is the most widely used, computationally straightforward, and flexible for comparison between various methods (Tong & Koller 2001; Settles 2009).

[2]This is a limitation of grammar-based approaches to similar problems. Since our method adapts and learns as we hand-code more and more uncertain observations, it begins to cover the realm of possible mismatches more broadly. Given the complexity of Congressional language and the continuing evolution of the legislative agenda, we are skeptical of the long-term returns to structure-based approaches seen specifically in Vannoni *et al.* (2019).

This paper uses a convolutional neural network (CNN) with a multi-layer perceptron (MLP) as our primary classifier. We note that a variety of other supervised learning methods can accommodate active learning, and the choice of model is, in general, the researcher's preference. We use uncertainty sampling, querying those examples with logits closest to 0.5.

Convolutional Neural Networks for Text Classification

We follow the example of Kim (2014) who defined a straightforward CNN for text classification as well as Zhang & Wallace (2015), who provide practical guidance on using such models. The primary advantages of using deep learning for our text analysis — rather than "bags of words" approaches (see those discussed in Grimmer & Stewart 2013) — are to model the similarity between words and to account for word order in text sequences. Both of these advantages involve considering words in context.

Underpinning the use of CNNs for text analysis is the distributed representation of words, wherein each word in a vocabulary is associated with a real-valued feature vector (Bengio *et al.* 2003; Mikolov *et al.* 2013; Pennington *et al.* 2014). These expressive vector representations encode many linguistic regularities and patterns, such as the relationships between synonymous words, and their use has been shown to improve the accuracy of supervised NLP tasks (Turian *et al.* 2010). Unsupervised training of word embeddings is typically accomplished by predicting the incidence of words given local context words, but practitioners may wish to instead randomly initialize word embeddings to be fully learned as parameters of their specific task (Rodriguez & Spirling 2021).

Concurrently with the maturation of distributed representations of words, convolutional neural networks have been shown to leverage word vectors for text classification effectively Zhang & Wallace (2015). The architecture adapted in this paper is that of Kim (2014), with practical guidelines for use outlined by Zhang & Wallace (2015). Prior to the convolutional network, the word tokens of the text to be classified are transformed to their real vector space word embeddings via a lookup table. Let d be the dimension of the word embeddings, and let $\mathbf{w}_i \in \mathbb{R}^d$ be the word embedding of the i-th word in the text. Each bill section is padded to be the same length n, and a single section is then represented as the concatenation (stacking) of the word embeddings that comprise it. We will denote the document (i.e., bill section) matrix as $\mathbf{W} \in \mathbb{R}^{n \times d}$ - the stacking of the word embeddings - with $\mathbf{W}_{i:j}$ denoting the sub-matrix of $\mathbf{W}$ from row i to row j.

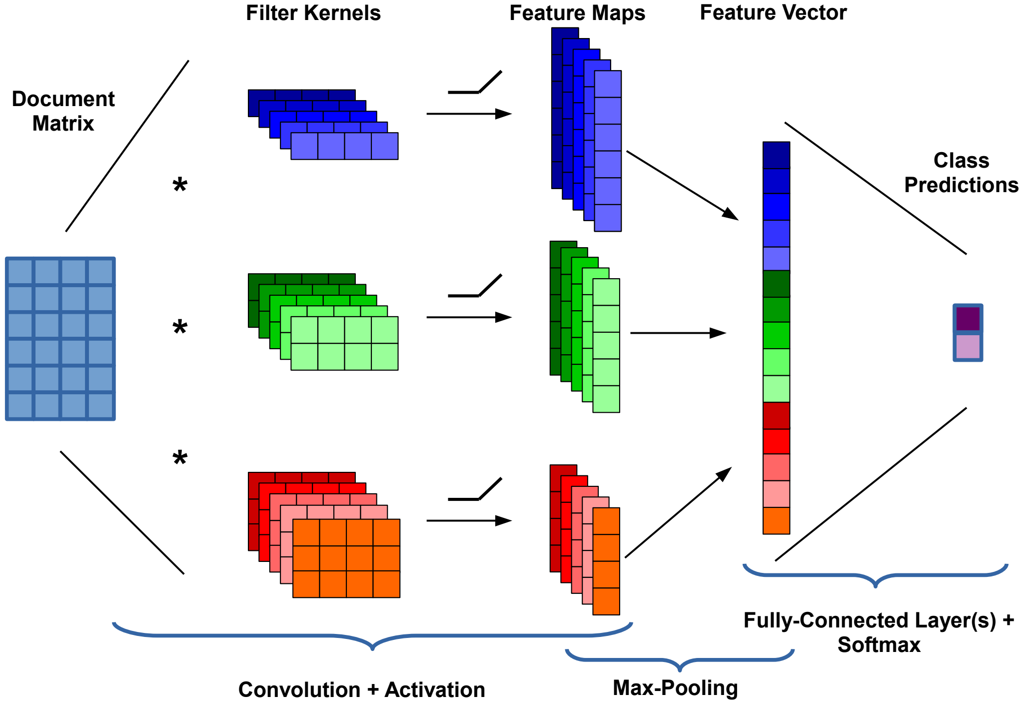
A convolutional filter is parameterized by a matrix $\mathbf{q} \in \mathbb{R}^{h \times d}$, where h is the region size of the filter. For text applications, h indicates how many words the filter operates upon at once (e.g., $h = 1$ corresponds to a single word, $h = 2$ to a bigram, and $h = n$ to an n-gram). Convolution is performed by applying the filter $\mathbf{q}$ to a window of words $\mathbf{W}_{i:i+h-1}$, which is accomplished via the summation of element-wise multiplication of the matrices, which shall be denoted by the $\cdot$ operator. The feature, c_i extracted by the operation is obtained by adding a bias term $b \in \mathbb{R}$ and applying a non-linear activation function f:

$$c_i = f(\mathbf{q} \cdot \mathbf{W}_{i:i+h-1} + b), \tag{1}$$

For our purposes, the non-linear activation function chosen is the *rectified linear unit* (ReLU - Glorot *et al.* (2011)), which is defined as: $f(x) = \max(0, x)$. The filter is applied to each possible contiguous window of h words in the matrix $\mathbf{W}$ to produce a feature map $\mathbf{c}$, a vector of features extracted by the operation described in equation (1). To increase the model's ability to capture relevant information from the text, multiple filters of the same size are used, with the idea being that they will extract complementary features from the same regions of text. Additionally, multiple filter sizes may be used within the same model.

Given our use of filters of different sizes and the varying lengths of text, a pooling scheme is used over the acquired feature maps to assemble a fixed-length feature vector for the text. Following Zhang & Wallace (2015), we choose 1-max-pooling, in which only the maximum activation from each feature map is retained and all such scalar values are concatenated to form the final feature vector. This strategy adheres to the intuition of choosing the "most important" feature from each map, and is furthermore computationally efficient (Zhang &

Figure 1: Illustration of CNN for text classification



For this example, $d = 4$, there are 5 filters of lengths {1,2,3}, and there are 2 classes.

Wallace 2015).

The above describes the process whereby a fixed-dimensional feature vector $\mathbf{c} \in \mathbb{R}^F$ may be obtained from text using a convolutional neural network, where F is the total number of convolutional filters in the network. This vector may be used for many tasks that can leverage a compressed representation of the text. For text classification, we use a *multi-layer perceptron* (MLP) to map the text feature vector $\mathbf{c}$ to a vector of scores for each possible class $\mathbf{s} \in \mathbb{R}^C$, where C is the number of possible categories. Our MLP comprises two fully-connected layers with ReLU nonlinearities. Each fully-connected layer applies a weight matrix $\mathbf{M}_{FC}$ to its input and adds a bias b_{FC} before applying the nonlinear activation function:

$$\mathbf{y} = f(\mathbf{M}_{FC}\ \mathbf{c} + b_{FC}) \tag{2}$$

Finally, we use the softmax function to compute the model probabilities for each class from the preceding scores. The model is trained by minimizing the binary cross-entropy loss between model predictions and true class labels with respect to authority delegation. Figure 1 depicts an illustration of the described architecture for binary classification of text. The architecture can accommodate arbitrarily many classes. [3]

Data

We utilize data for all versions of all bills (both successful and unsuccessful) from the 110$^{\text{th}}$ and 111$^{\text{th}}$ Congresses. We separate each version of each bill into titles and analyze them at the bill section level. We do this for three reasons: first, because each title in each bill deals with a particular agency or activity and likely contains an entire delegatory phrase, keeping the task more straightforward. Second, any given bill could delegate authority to multiple agencies in multiple titles, so to avoid missing any additional delegations, we wanted to reduce it to units that are about a single delegation. The final reason we chose bill titles is that bill titles are the smallest comparable distinct units of a bill: titles are more comparable to one another than either a sentence or an entire bill would be.

While it might seem advantageous to go even more fine-grained than the bill section, there are fundamental limitations drawn from how bills are written. For example, the most common unit of analysis in natural language processing tasks is the sentence. The construction of any given sentence within a bill is, however, is contingent primarily on the remaining information presented at the title level. Furthermore, in titles where authority is delegated to an agency, sentences may be less precise than the section, which provides much-needed context. Studying bills at the title level will ultimately let us make better inferences about

[3] For the results presented in this paper, we use the CNN architecture described here with 64 filters each of sizes $\{1, 2, 3, 4, 5\}$. We choose a word embedding dimension of $d = 300$, and our dataset yields a vocabulary size of 5775 words. Our first MLP hidden layer has 64 neurons, while the second has 32. To train our model, we optimize the parameters using the Adam optimizer (Kingma & Ba 2014) with a learning rate of 0.0001 and batch size of 64 bill sections. The loss to minimize is the binary cross-entropy loss, with bill sections labeled according to whether they delegate authority or not. We allow the model to train for 13 epochs and regularize the model using Dropout (Srivastava *et al.* 2014) in the MLP with a drop-rate of 15%.

the agencies or programs to which Congress has delegated.

Our complete dataset has several components. The division of our dataset is into labeled and unlabeled sub-datasets. The labeled component comprises 2098 bill sections from the 110^{th} that were read by human annotators and assigned a binary label with respect to delegation; the unlabeled component contains the remaining 137,616 bill sections from the 110^{th} Congress as well as all sections from the 111^{th} Congress. As is standard practice within machine learning, we divided our labeled data into subsets for training, validation, and test. In performing this splitting, we ensured that all titles/versions from the same bill were apportioned to the same subset. We randomly divided the bills in our labeled data into training, validation, and test using proportions of 65%, 15%, and 20%, respectively. The remaining bills from the 110^{th} Congress were also apportioned into training, validation, and test using the same proportions. The numbers of bills (by bill number) and examples (e.g., bill section) for each subset are presented in the Appendix.

Delegation Coding

Essential to our project is a consistent definition of delegation to administrative agencies. An act of delegation is a mandate or permission for a federal agency or program (including the President) to exercise public authority in some way (see McCubbins *et al.* 1987; Kiewiet & McCubbins 1991; Huber & Shipan 2002; Gailmard & Patty 2012, for a discussion of this point). For our task, we stated that allocating money for federal agencies to spend, instructing agencies to promulgate rules, granting agencies the ability to exempt themselves from preexisting rules, requiring agencies to compile reports or commission pilot studies, and charging agencies with the enforcement of specific policies are all delegation (Kiewiet & McCubbins 1991).

For the hand-coding, we gave straightforward instructions as to how we identify delegation. First, is Congress acting upon an administrative agency? This will include all references to both the agency itself and the person in charge of that agency. We operated with a list of administrative agencies and matched each instance of delegation to one of those agencies.[4]

Second, what is the title asking the agency to do? In general, Congress delegates authority by asking an agency to perform a specific task, collect information, write new regulations, hire people, write a report to Congress on their activities, delegate authority to sub-agencies or outside of government, and make or distribute an award, among many other things. A bill is not delegating authority if it only appropriates money, if they are referencing actions already taken, or if Congress is writing new rules or regulations. Keeping these actions separate allows us to track statutorily derived authority for the agencies, not merely what funds they have been allotted.

Below are example bill titles that the active learner selected in early runs as uncertain and how they were coded.

- `Section.2402. energy conservation projects. using amounts appropriated pursuant to the authorization of appropriations in section.2403.a.6, the secretary of defense may carry out energy conservation projects under chapter 173 of title 10, united states code, in the amount of 800000` (delegates authority to an administrative agency)

- `Section.2. reemployment of foreign service annuitants... the authority of the secretary to waive the application of subsections a through d for an annuitant pursuant to subparagraph c of paragraph 1 shall terminate on September 30, 2008. the authority of the secretary to waive the application of subsections a through d for an annuitant pursuant to subparagraph c ii of paragraph 1 shall terminate on September 30, 2009` (does NOT delegate authority to an administrative agency)

In the above examples, it is clear why the algorithm would have selected them as ambiguous classifications and why human readers would classify them correctly. Take the top section, dealing with "Energy Conservation Projects." A quick read of the section makes

[4] Most often, if Congress is referring to a governmental entity (except for organizations already within Congress, which they make apparent), it is an administrative agency. We exception for delegating to the courts or states and local governments because those tasks are defined differently.

it clear that Congress is delegating authority to Defense (through the Secretary of Defense) to spend $800,000 on energy conservation projects. The classifier may have had issues with the added verbiage of the task (may carry out) and the addition of US code language in between. In the second section, where Congress is setting up "foreign service annuitants", it is clear that the agency is not being given an extra task or authority but only describes how applications must be processed (and when they terminate). This section is an example where the language (context-free) would indicate the possibility of delegating authority, but the additional context (plus a close reading) makes it clear that this is not occurring. These are only two examples pulled from an early run of the active learning module, set up to illustrate the nature of the classification task.

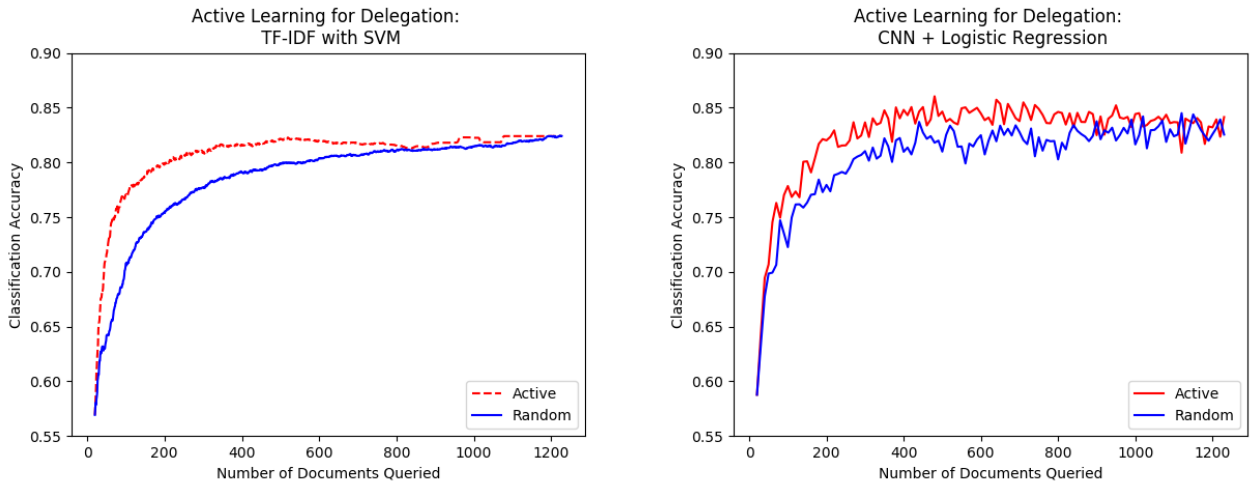
Model Performance

We propose two aspects regarding our convolutional neural network and active learning model performance. First, we suspect the CNN will outperform a traditional word-document-matrix vector-space method of text classification. For comparison, we use baselines of a term-frequency, inverse-document-frequency (tf-idf) text representation with linear support vector machine (SVM), L1-penalized logistic regression (LASSO), and random forest classifiers.[5] Second, we expect that incorporating active learning into the classifiers will outperform random sampling as additional documents are appended to the training data.

To evaluate active learning performance, we employed a standard demonstration. Generally, our scheme involves artificially restricting the training set to a small number of examples and iteratively evaluating performance on the validation set while augmenting the training set. When adding to the training set, we use either active learning or random sampling for selecting new examples before retraining the model and then re-evaluating performance. As the training set grows, we expect validation performance to increase, but more

[5] See Appendix Section 3 for a formal description of the baseline models and Miller *et al.* (2019) for work using tf-idf with active learning.

Figure 2: Active learning querying versus random sampling for SVM and CNN models.



Note that the active learning models generally outperforms the random sampled models and that the CNN generally outperforms the SVM.

rapidly for the active learner, as it has judiciously queried examples to strengthen performance. We begin by randomly selecting ten documents from our training set. We then query ten additional documents from training, either actively or randomly, and retrain the model before evaluating again. This process continues until the whole of the training set has been queried.

Figure 2 shows this active learning demonstration for our neural model and for the SVM baseline. In both cases, the active learner outperforms the random sampler in that classification accuracy increases more dramatically as training documents are added. The trend is particularly apparent for training sizes of 20-800 examples for the SVM and 200-600 examples for the CNN. As expected, when the entire training set is used, performance converges between active and random sampling. We note that the CNN's active/random sampling curves exhibit significantly more variability than for the SVM due to the inherent stochasticity in neural architectures, which are more subject to the randomness of parameter initializations. To mitigate this and the randomness of sampling training examples, we obtained these active learning curves by averaging the results over five trials.

Also apparent in Figure 2 is that the CNN outperforms the SVM on classification performance. In fact, the CNN outperforms all three of our tf-idf baselines: SVM, LASSO, and Random Forest. In Table 1, we present the classification accuracies on both the validation and test dataset splits for all baseline classifiers and the neural model. We additionally used the fully-trained CNN to actively query an additional 200 examples from the unlabeled training partition. After incorporating those new examples into our dataset and retraining, we obtain the "Post-Query" accuracies in Table 1, where modest improvement is evident.

Table 1: Delegation Classification Accuracy

		Pre-Query		Post-Query	
Features	Classifier	Val.	Test	Val.	Test
TF-IDF	SVM	82.4	87.1	82.6	86.3
	Random Forest	81.2	87.3	82.4	88.4
	LASSO	83.9	87.9	85.9	87.1
CNN	MLP	**86.5**	**90.2**	**87.6**	**90.4**

Classification accuracy for tf-idf baseline models and our neural network – convolutional neural network with a multilayer perceptron (CNN and MLP). We provide accuracies on both the validation and test splits, as well as before and after the addition of new documents queried by the active learner.

Classification and Delegation Results

With delegation predictions for each version of each bill section from the 110th and 111th Congress, the remainder of the paper will examine the consequences of these delegation classifications. First, we examine how closely our classification of delegatory sections concurs with existing measures of discretion (in particular, we validate the textual measure popularized by Huber & Shipan (2002)). Second, we identify aspects of Congressional lawmaking which predict delegatory load for each bill; we compare this to results from Epstein & O'Halloran (1999) but focus mainly on validating results with our measure against existing theories of lawmaking. Furthermore, third, we identify which agencies receive the bulk of delegated authority in Congress. Finally, we directly compare the effects of divided government in the 110th Congress with unified during the 111th.

In this section, we both show that our measure of delegation is consistent with extant theories on delegation and lawmaking, but also that some of the most widely used measures of delegation overlook some of the points a section-by-section delegation coded dataset provides, especially when contrasted with measures that are only used on "significant statutes." We also examine how the legislative process changes the use of delegatory language and the first to test delegation models on bills throughout the process, including those that do not pass.

What does Delegation Look Like?

First, we compare our measure of delegation to the most widely used proxy–bill length–which was first introduced in Huber & Shipan (2002). The logic of the measure is straightforward; Huber & Shipan (2002) claim that longer bills are more likely to require specific returns from agencies, and this generally entails more fine-grained oversight and control. Shorter bills, ones that discuss delegation and discretion more broadly, are less likely to have the same requirements and instead reflect less deference from the legislature. It has been accepted by the literature (see Clinton *et al.* (2012) as an example) that longer bills mean more delegation.

If Huber & Shipan (2002)'s measure of delegation is accurate, we would expect the length of a given section to predict delegation, which we can compare directly to the predictions from our classification model. We can also examine discretion for the aggregate of a bill over all of its sections using the delegation ratio (Epstein & O'Halloran 1999; Anastasopoulos & Bertelli 2020). Another widely used general measure of discretion, the delegation ratio, is defined as the total number of sections delegating authority divided by the total number of sections; essentially, how much of the bill delegates authority. This penalizes omnibus legislating and other forms of massive legislation that attempt to do many things at once since raw counts would overweight these bills.

Table 2 compares delegation measures for the 110$^{\text{th}}$ Congress. First, we observe that the number of words corresponds with the likelihood a given bill section delegates authority (Models 1 & 2). However, total discretion in a bill — as measured by the delegation ratio —

Table 2: Comparing Total Words to Delegation Measures for 110$^{\text{th}}$ Congress

	Delegation		**Delegation ratio**	
	Logit	**Logit w/ Mixed-Effects**	**Beta**	**Beta w/ Mixed-Effects**
	Model 1	**Model 2**	**Model 3**	**Model 4**
Words/1000	0.829*	0.942*	−0.002*	−0.001*
	(0.008)	(0.006)	(0.0003)	(0.0003)
Constant	−1.087*	−1.584*	0.002	0.132*
	(0.007)	(0.009)	(0.010)	(0.021)
N	139714	139714	8847	8847
Log Likelihood	−82265.290	−77319.430	754.258	876.266

* = $p < .05$

is not associated with the total number of words (Models 3 & 4). Importantly, the delegation ratio is negatively associated with total bill length, a relationship that directly challenges the efficacy of bill length as a proxy for discretion. The relationship between the length of bills is only weakly associated with how much agency discretion there is and that indirect measures of agency discretion may have been confounding, even though this measure is pretty widely used (see Huber & Shipan 2002; Clinton *et al.* 2012, as examples though there are many more as well). This demonstrates the value of using a text-based classification scheme of delegation section-by-section.

We see this further illustrated in Table 3, which shows the confusion matrix for Model 1 from Table 2 (which used only bill section length) against our model predictions. Overall, predictive accuracy is middling, around 71.2%, which is lower than our test set predictive rate of 90% from Table 1. Considering the dramatic difference in model sophistication, the lack of agreement between the models is unsurprising.

Table 3: Confusion Matrix for Word Length Measure to Machine Labeled Sections

	Label from Deep Learning:	
Label from Section Length	Does Not Delegate	Delegates
Does Not Delegate	84194	36312
Delegates	4892	14316

We see that relying only on word count, a logit model predicts non-delegatory sections in agreement with our model at 94.1%. However, for sections that delegate authority, the logit model performs significantly worse: 14,316/50,628 correct, only 28.3% agreement with our predictions. In terms of predictive modeling, knowing the total number of words helps eliminate the non-delegatory sections, but is *worse than random* for longer bill sections. This is consistent with total word length for a section being a poor proxy for overall delegation and deference to an agency. Similarly, if we use median or mean words selection criteria (no model, all sections above the mean/median are coded as delegatory), we find that predictive accuracy stays about the same, at 72.1% and 69.4% accuracy, respectively. This suggests that using a word count/volume proxy for delegation or discretion provides only a weak signal: the performance of the logistic regression suggests even more strongly that these designations are almost certainly missing entire types of delegatory action, in most likely systematic ways.

Combining this with the discussion of the delegation ratio models from Table 2, we see that though bill length provides some information consistent with discretion, there are many issues with using it directly. Longer bills almost certainly delegate more frequently than shorter bills but is that a function of the type of discretion they are providing, or is it a tautology that *longer bills tend to do more*? With the rise of omnibus bills and the pervasiveness of hitchhiker legislating (see Krutz 2001; Casas *et al.* 2020, for more), simply identifying that these longer bills delegate more is insufficient.[6]

What Bills Delegate?

Next, we examine what differentiates bills that delegate from bills that do not and how delegation in a bill is associated with various legislative institutions. For the remaining analysis in this section, we aggregate the delegating activities at the bill version level. We

[6] Each of the following results in this paper are presented with bill length as a robustness check in Appendix Section 8, with only one result differing.

include data for both the 110^{th} and 111^{th} Congresses.[7] We have multiple versions of each bill as it progresses through the legislative process, but aggregate delegation separately for each unique version of the bill.[8] We count the sections that delegate authority to administrative agencies as our primary variable of interest here.[9]

To make the analysis comparable to the delegation ratio, but not as constrained by forcing the results to be between 0 and 1, we include the total number of bill sections as a covariate in each model. Because only around 60% of bills delegate authority, we use zero-inflated negative binomial (ZINB) models.[10] A ZINB also assumes that the data generating process for zeros is distinct and separable from the data generating process for bills that delegate; most zero-delegation bills are themselves inconsequential legislation or are appropriations bills with no extra discussion of agency control. Because those bills are fundamentally different from legislation that delegates, we need to account for this two-stage selection process directly, which is what a ZINB does.[11]

We included as predictors several variables from the literature that should predict the size and scope of each bill's number of delegating sections. First, we include the number of committees a bill is referred to. Next, we include information about the bill sponsor: whether or not they are a committee chair or subcommittee chair and their party. We then include variables about bill progression: a dummy for whether or not the bill is reported out of committee and another if the bill passes the chamber. These models are both exploratory and confirmatory since we believe this to be a good model of bill delegation; we would expect each element that makes a bill more encompassing (multiple committee referrals, having the

[7] We did not include the 111^{th} in the first set of analyses in this section so we could validate our measure against length directly, factoring both hand-coded sections and machine-coded sections, though results using both Congresses remained constant.

[8] Each version of the bill is represented as its own unique bill. For example, if a bill was introduced in the House, referred from a committee, passed the House, passed the Senate, and signed into law, we would represent each of these stages as different versions of the same bill.

[9] Descriptions of the variables used below, including our measure of delegation, are available in Section 5 of the Appendix.

[10] See Appendix Table 6 using the delegation ratio with a beta regression as a robustness check.

[11] We also examine the same sets of models by excluding purely commemorative legislation in Appendix Table 7 and nothing of note changes.

chairs champion the bill) would correspond with additional delegation. This should confirm with the political/institutional accounts of delegation first analyzed empirically in Epstein & O'Halloran (1999), but also mindful of process concerns.

Finally, we include a dummy for the term: specifically, if it is the 111$^{\text{th}}$ Congress (and therefore a unified Democratic House, Senate, and Presidency). Similar to Lowande (2018), we exploit the fact that the transition from the 110$^{\text{th}}$ Congress to the 111$^{\text{th}}$ Congress moves from a Democratic-controlled House and Senate, but Republican president, to unified Democratic control. Since the Congress side remains stable, we can assume that major changes in delegation and strategic choices about bill writing are largely a function of responsiveness (or non-responsiveness) to the changing partisan conditions of the presidency (Farhang & Yaver 2016).

We obtain a clear picture of delegation in Congress as a complex policy process that is altered significantly by Congressional leadership and strategic decisions made throughout the legislative process. To address this fully, we use four different models that test various elements of this process but use different independent variables. Model 1 only includes the number of committees a bill is referred to, length, and a dummy for unified or divided government. Model 2 adds sponsorship effects. Model 3 adds bill progression effects. Model 4 only includes bills that became law — excluding the process variables.[12]

An advantage of the ZINB model is that it separates inquiry into what makes a bill more likely to delegate at all, versus questions about the scope of the delegation, in which both institutional and process variables seem to matter a lot. Consistent with Epstein & O'Halloran (1999), bills that are referred to a larger number of committees do seem more likely to contain more delegatory sections, but more referrals are not associated with a greater propensity to delegate in the first place. The sponsorship effects are consistent with this: having a sponsor chair a committee or subcommittee that the bill is referred to increases

[12]These results are robust to excluding commemorative bills and to focusing on only significant statutes, as defined in the Congressional Bills project and Mayhew (1991) respectively, and are available in Section 8 of the Appendix.

Table 4: Zero-Inflated Negative Binomial Model of Number of Delegating Sections

	All Bill Versions			**Laws**
	Model 1	**Model 2**	**Model 3**	**Model 4**
Number of Delegating Sections: Negative Binomial				
Number of Referrals	0.122*	0.119*	0.143*	0.101*
	(0.009)	(0.009)	(0.009)	(0.046)
Number of Bill Sections	0.032*	0.030*	0.030*	0.012*
	(0.0005)	(0.0005)	(0.0005)	(0.001)
Unified Gov?	−0.059*	−0.067*	−0.066*	−0.145
	(0.017)	(0.017)	(0.017)	(0.122)
Sponsor Chair of Committee		0.206*	0.158*	0.443*
		(0.026)	(0.027)	(0.141)
Sponsor Chair of Subcommittee		0.127*	0.098*	0.327*
		(0.025)	(0.025)	(0.149)
Republican		−0.181*	−0.196*	−0.391
		(0.021)	(0.021)	(0.212)
Report out of Committee			0.354*	
			(0.023)	
Pass Chamber			−0.336*	
			(0.023)	
Delegation: Logit				
Number of Referrals	0.042	0.040	−0.389*	0.249
	(0.051)	(0.052)	(0.109)	(0.230)
Number of Bill Sections	−1.618*	−1.591*	−1.592*	−1.375*
	(0.055)	(0.054)	(0.055)	(0.202)
Unified Gov?	−0.107*	−0.086	−0.070	−0.665*
	(0.060)	(0.061)	(0.062)	(0.330)
Sponsor Chair of Committee		0.560*	0.430*	0.318
		(0.138)	(0.142)	(0.479)
Sponsor Chair of Subcommittee		0.068	0.037	−0.351
		(0.119)	(0.120)	(0.456)
Republican		0.183*	0.168*	−0.348
		(0.067)	(0.068)	(0.393)
Report out of Committee			0.105	
			(0.094)	
Pass Chamber			0.627*	
			(0.116)	
N	28907	28907	28907	816

* = $p < .05$

the delegatory load considerably, suggesting something about how the parties execute their agendas. Having chairs sponsor ambitious bills would be consistent with parties devoting additional attention and energy to these types of legislative problems. The fact that this effect is stronger on bills that ultimately become law (Model 4) is consistent with a top-down view of the lawmaking process, with delegation to key actors (the chairs) playing an essential role in shepherding through the most impactful legislation.

The final part of the model examines the expansion of legislation as it succeeds in going through the committee process but then shrinking to pass the chamber. Since the committee process is dominated by majority party gatekeeping, it is unsurprising that the legislation with the widest delegatory scope would emerge from that stage, only to shrink down once it hits the floor and has to pass the chamber. This suggests negotiation is a process of narrowing policy concerns, consistent with the literature regarding legislative winnowing (Krutz 2005; Cox & McCubbins 2005).

The final result is a bit unexpected. The consistent story from the literature on delegation and governance is that delegation is more likely to happen, and is broader, during times of unified rather than divided government, consistent with the "ally principal" (Moe 2012; Farhang & Yaver 2016). However, those studies only look at "significant" statutes, those designated by Mayhew (1991). Since we consider all versions of all legislation, it is not surprising that there are differences. Indeed, as we will see when we look at the agencies receiving delegation, there is strong evidence that the very extreme end of legislation may indeed be getting more broad during unified government. But if we think of delegation in terms of all legislation, "significant" or not, we find no such story.[13] This is consistent with Shaffer (2020), who finds similar discrepancies in characteristics of legislation between "significant laws" and all laws; he finds that expected differences in scope between divided and unified government are mostly an artifact of the use of significant laws. This is also consistent with the argument from Kiewiet & McCubbins (1991) and McNollgast (1999), who see the division between Congress and the Presidency as less crucial than the division between the House and Senate; since both the 110th and 111th Congresses are controlled by Democrats, we cannot test whether this arrangement matters more. It is, however, a

[13] This use of Mayhew's significant statutes is not an unusual feature of the literature; since Epstein & O'Halloran (1999), it is common in studies of delegation to focus only on these bills (Farhang & Yaver 2016). Since most of what we find on delegation is consistent with Epstein & O'Halloran (1999), this distinction does not change most of the results. For the unified/divided question, there is a difference.

plausible explanation for the observed patterns.[14]

What Agencies Get Delegated To?

Next, we investigate to whom authority is delegated. There is some disagreement over whether delegation is primarily ideological (i.e., delegate to friendly agencies), political (delegate more broadly to friendly administrations), or transactional (delegate to agencies to address specific policy concerns). There is mixed evidence for all three theories. This is a complicated measurement problem because assessing the ideology of agencies is challenging. Focusing on agency heads, Bonica *et al.* (2015) use campaign contributions to assess how the political appointees at the tops of agencies change in terms of ideology over time. This would be key if delegation is a strictly political question since the heads of agencies can be thought of as a direct representation of their administration's policies. A different way of conceptualizing this is to focus on the federal civil service and assess the politics of the bureaucrats employed in each agency (Clinton & Lewis 2008; Richardson *et al.* 2018).

We attempt to discern the delegatory approach by Congress by observing the delegatory patterns in our new dataset. The transition from the Bush Administration to the Obama administration denotes a clear opportunity to view the preferences of Congress remaining stable (Democratic House and Senate) while the White House changes partisan affiliation. If there are substantial shifts in delegation in divided versus unified government, this would be strong evidence that MCs are treating the agencies as the sum of their political appointments; if there are minimal or no changes, this would be evidence that MCs treat the agencies as stable reflections of the bureaucracy, and are more consistent with the administrative state (Lowande 2018).

Figure 3 shows the proportion of delegation made to cabinet-level agencies by party for the 110th and 111th Congresses respectively. We do not find a clear ideological picture be-

[14]When we run simplified versions of our models on just Mayhew's significant laws, we find small but positive effects of unified government. Since there were only 26 significant bills in these Congresses, it is unsurprising that these effects were not statistically significant.

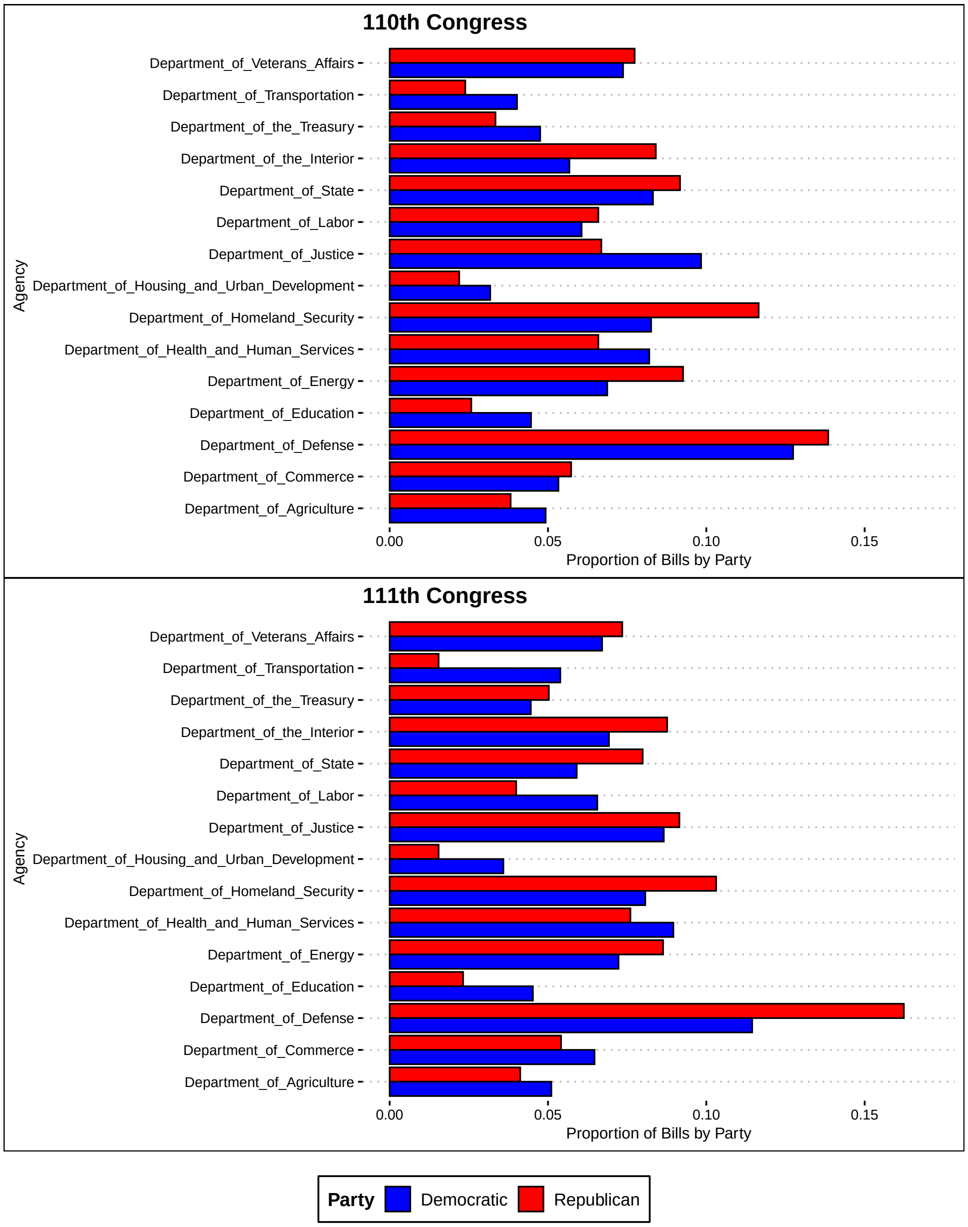


Figure 3: Proportion of Bills by Party Delegating to Cabinet Level Agencies

tween both Congresses; there are some expected results from agencies that are more political in nature, but the contrast is not nearly as stark as the agency ideology literature would have suggested. This is consistent with Lowande (2018), who also studies the same Congresses we do, for a plausible account that shows that changes in agency ideology matter less to Congress in terms of Congressional oversight than do interpersonal connections. This suggests that in purely statutory terms, formal delegation is less driven by ideological interests than by particularist interest—hence the Republican delegation to the Department of the Interior, which largely consists of references to the National Park Services and other agencies that are mostly non-ideological and instead are most likely receiving attention because of constituent characteristics and concerns.

We next turn to modeling total agencies delegated to by Congress. We limit our data to bills that delegate authority. Our DV is the number of unique agencies mentioned in a bill that delegates authority; this can be viewed as an alternative way of measuring the scope of delegation, covering the scope of agencies impacted by a given bill rather than the total number of delegating sections.[15] These bills, by definition, cover a larger range of policy space as the number of agencies increases, reflecting a style of legislating that is more indicative of "omnibus" lawmaking rather than singular legislation one at a time. To model this, we use a negative binomial regression on the number of agencies included. The independent variables included are the same as they were for Table 4.

We see results in Table 5. Consistent again with Epstein & O'Halloran (1999), we find delegatory bills cover a larger number of agencies when a bill is referred to multiple committees. Also consistent with our earlier findings, we see sponsor effects with the committee and subcommittee chairs and a negative effect of being a Republican (though we cannot separate Republican-specific effects from minority party effects). We also see the same pattern for the number of agencies as we do for delegatory sections in the process variables; bills that report out of committee cover more agencies (and delegate more), but bills that pass the chamber lose agencies mentioned and delegatory sections. This is consistent with the committee process expanding the scope of legislation, while floor negotiations drive the scope of bills down, which follows the logic of the increased efficacy of negative bargaining (the act of dropping controversial provisions) rather than positive bargaining—both corollaries to theories of positive and negative agenda control (see Aldrich & Rohde 2000; Cox & McCubbins

[15] We use the list of agencies from Richardson *et al.* (2018).

Table 5: Negative Binomial Model of Number of Agencies in Delegating Bills.

	DV: **Number of Agencies**			
	All Bill Versions			**Laws**
	Model 1	**Model 2**	**Model 3**	**Model 4**
Number of Referrals	0.036*	0.038*	0.040*	0.047
	(0.008)	(0.008)	(0.008)	(0.041)
log(Number of Bill Sections)	0.733*	0.692*	0.688*	0.784*
	(0.007)	(0.008)	(0.008)	(0.043)
Unified Gov?	0.096*	0.090*	0.095*	0.114
	(0.018)	(0.018)	(0.018)	(0.118)
Sponsor Chair of Committee		0.188*	0.155*	0.333*
		(0.026)	(0.027)	(0.138)
Sponsor Chair of Subcommittee		0.268*	0.241*	0.070
		(0.025)	(0.025)	(0.147)
Republican		−0.071*	−0.066*	−0.478*
		(0.023)	(0.023)	(0.231)
Report out of Committee			0.151*	
			(0.023)	
Pass Chamber			−0.051*	
			(0.024)	
N	16629	16629	16629	333

$^* = p < .05$

2005).

The most substantial difference between the agency scope models of Table 5 and the overall delegatory scope models of Table 4 is for the unified government variable. For agency scope, unified government is strongly and positively associated with bills delegating to more agencies, whereas for total delegatory scope, unified government is negatively associated with bills having a greater delegatory scope. This result suggests that the overall breadth of bills pursued under unified government may be greater than under divided government, but only in that they cover a great range of lawmaking areas and, therefore, impact more agencies — a finding that extends Farhang & Yaver (2016)'s beyond significant legislation, and may explain the discrepancies. Changes in scope also could reflect an over-reliance in recent Congresses on omnibus legislating, making "hitchhiking" legislative strategies more common (Casas *et al.* 2020).

Discussion

In this paper, we have demonstrated how an active learning convolutional neural network for classifying text can be used to study a complex problem in political science: agency delegation from Congress. We hope that the methods we have proposed are clear and usable for other applications and that gains in classification accuracy, while reducing the need for extra documents to be hand-labeled, helps researchers tackle challenging classification problems. This information is part of an ongoing endeavor to learn how Congress uses statutory language to enact its agenda and how the modern legislative process provides oversight and guidance to implement policies. These classifications will help provide the means through which we can test our theories of delegation and Congressional oversight on a larger scale basis and provide the nexus for increased research into the implications of statutory language.

After classifying bill sections, we used this newly labeled data in multiple ways. First, we compared our measure of delegation with a well-known proxy. Second, we analyzed delegation throughout the legislative process. Third, we examined the range of agencies delegated to. Finally, we compared delegation under unified and divided government. We added answers to questions that had previously only been partially answered in the literature, and we extended these findings to bills along the legislative process and all legislation generally. Moving further, we can expect our data to help illuminate different claims about how and when Congress changes delegating powers. Exploiting the transition between the Bush and Obama administrations between the 110^{th} and 111^{th} Congresses gives us ample opportunities to more precisely examine how delegatory strategies evolve as partisan conditions change. Although we did explore how much authority is delegated and the number of agencies delegated to, more precise causal estimands are possible, especially in exploiting the timing specification in delegation. More direct tests of the ideological characteristics of delegation would be a fruitful extension.

An additional next step would be to compare our data on delegation with data on oversight hearings and see empirically how intertwined these two tools of Congressional engagement with the administrative state are. Linking formal delegations with strategic oversight hearings would enable us to understand how Congress oversees the authority it tasks agencies with and could shed more light on what it means to delegate. An additional extension on this project would be to link these formal delegations to administrative agencies with appropriations riders and other funding sources from Congress and see if and when Congress also provides the resources to perform the tasks it lays on the bureaucracy.

We expect that deep and active learning models of text classification can be used in various ways beyond agency delegation. In particular, we believe that problems involving the classification of complex texts, particularly those requiring expertise for labeling, can benefit from schemes similar to our method. We envision that this is only the first of many applications of this model to legislative texts. We are now interested in performing the same task on legislative constraints and regulations next to get a fuller sense of delegation in context, which opens up the scope of legislative outcomes dramatically, and encourages a clearer thinking about constructs.

References

Aldrich, John H, & Rohde, David W. 2000. The consequences of party organization in the House: The role of the majority and minority parties in conditional party government. *Pages 31–72 of: Polarized politics: Congress and the president in a partisan era.*

Anastasopoulos, L Jason, & Bertelli, Anthony M. 2020. Understanding delegation through machine learning: A method and application to the European Union. *American Political Science Review*, **114**(1), 291–301.

Bengio, Yoshua, Ducharme, Rejean, Vincent, Pascal, & Janvin, Christian. 2003. A Neural Probabilistic Language Model. *The Journal of Machine Learning Research.*

Bonica, Adam, Chen, Jowei, & Johnson, Tim. 2015. Senate Gate-Keeping, Presidential Staffing of 'Inferior Offices,'and the Ideological Composition of Appointments to the Public Bureaucracy. *Quarterly Journal of Political Science*, **10**(1), 5–40.

Casas, Andreu, Denny, Matthew J, & Wilkerson, John. 2020. More effective than we thought: Accounting for legislative hitchhikers reveals a more inclusive and productive lawmaking process. *American Journal of Political Science*, **64**(1), 5–18.

Clinton, Joshua D, & Lewis, David E. 2008. Expert opinion, agency characteristics, and agency preferences. *Political Analysis*, 3–20.

Clinton, Joshua D, Bertelli, Anthony, Grose, Christian R, Lewis, David E, & Nixon, David C. 2012. Separated powers in the United States: The ideology of agencies, presidents, and congress. *American Journal of Political Science*, **56**(2), 341–354.

Cox, Gary W, & McCubbins, Mathew D. 2005. *Setting the agenda: Responsible party government in the US House of Representatives*. Cambridge University Press.

Epstein, David, & O'Halloran, Sharyn. 1996. Divided government and the design of administrative procedures: A formal model and empirical test. *The Journal of Politics*, **58**(2), 373–397.

Epstein, David, & O'Halloran, Sharyn. 1999. *Delegating powers: A transaction cost politics approach to policy making under separate powers*. Cambridge University Press.

Farhang, Sean, & Yaver, Miranda. 2016. Divided government and the fragmentation of American law. *American Journal of Political Science*, **60**(2), 401–417.

Gailmard, Sean, & Patty, John W. 2012. Formal models of bureaucracy. *Annual Review of Political Science*, **15**, 353–377.

Glorot, Xavier, Bordes, Antoine, & Bengio, Yoshua. 2011. Deep Sparse Rectifier Neural

Networks. *In: 14th International Conference on Artificial Intelligence and Statistics (AISTATS).*

Grimmer, Justin, & Stewart, Brandon M. 2013. Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political analysis*, **21**(3), 267–297.

Huber, John D, & Shipan, Charles R. 2002. *Deliberate discretion?: The institutional foundations of bureaucratic autonomy.* Cambridge University Press.

Kiewiet, D Roderick, & McCubbins, Mathew D. 1991. *The logic of delegation.* University of Chicago Press.

Kim, Yoon. 2014. Convolutional Neural Networks for Sentence Classification. *Pages 1746–1751 of: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP).*

Kingma, Diederik P., & Ba, Jimmy. 2014. *Adam: A Method for Stochastic Optimization.*

Krutz, Glen S. 2001. *Hitching a ride: Omnibus legislating in the US Congress.* Ohio State University Press.

Krutz, Glen S. 2005. Issues and institutions: "Winnowing" in the US Congress. *American Journal of Political Science*, **49**(2), 313–326.

Lowande, Kenneth. 2018. Who Polices the Administrative State? *American Political Science Review*, **112**(4), 874–890.

Mayhew, David R. 1991. *Divided we govern.* Yale University New Haven, CT.

McCubbins, Mathew D. 1985. The legislative design of regulatory structure. *American Journal of Political Science*, 721–748.

McCubbins, Mathew D, Noll, Roger G, & Weingast, Barry R. 1987. Administrative procedures as instruments of political control. *Journal of Law, Economics, & Organization*, **3**(2), 243–277.

McCubbins, Matthew D, Noll, Roger G, & Weingast, Barry R. 1989. Structure and process, politics and policy: Administrative arrangements and the political control of agencies. *Virginia Law Review*, 431–482.

McNollgast. 1994. Legislative intent: The use of positive political theory in statutory interpretation. *Law & Contemp. Probs.*, **57**, 3.

McNollgast. 1999. The political origins of the Administrative Procedure Act. *Journal of Law, Economics, & Organization*, 180–217.

Mikolov, Tomas, Sutskever, Ilya, Chen, Kai, Corrado, Greg, & Dean, Jeffrey. 2013. Distributed Representations of Words and Phrases and their Compositionality. *Advances in Neural Information Processing Systems*.

Miller, Blake, Linder, Fridolin, & Mebane, Walter R. 2019. Active Learning Approaches for Labeling Text: Review and Assessment of the Performance of Active Learning Approaches. *Political Analysis*, 1–20.

Moe, Terry M. 2012. Delegation, control, and the study of public bureaucracy. *In: The Forum*, vol. 10. De Gruyter.

Pennington, Jeffrey, Socher, Richard, & Manning, Christopher D. 2014. GloVe: Global Vectors for Word Representation. *Empirical Methods in Natural Language Processing*.

Richardson, Mark D, Clinton, Joshua D, & Lewis, David E. 2018. Elite perceptions of agency ideology and workforce skill. *The Journal of Politics*, **80**(1), 303–308.

Rodriguez, Pedro, & Spirling, Arthur. 2021. Word Embeddings: What works, what doesn't, and how to tell the difference for applied research. *The Journal of Politics*.

Settles, Burr. 2009. *Active learning literature survey.* Tech. rept. University of Wisconsin-Madison Department of Computer Sciences.

Shaffer, Robert. 2020. Power in Text: Implementing Networks and Institutional Complexity in American Law. *Journal of Politics.*

Srivastava, Nitish, Hinton, Geoffrey, Krizhevsky, Alex, Sutskever, Ilya, & Salakhutdinov, Ruslan. 2014. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research*, **15**(56), 1929–1958.

Tong, Simon, & Koller, Daphne. 2001. Support vector machine active learning with applications to text classification. *Journal of machine learning research*, **2**(Nov), 45–66.

Turian, Joseph, Ratinov, Lev, & Bengio, Yoshua. 2010. Word Representations: A Simple and General Method for Semi-Supervised Learning. *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, July.

Vannoni, Matia, Ash, Elliott, & Morelli, Massimo. 2019. Measuring Discretion and Delegation in Legislative Texts: Methods and Application to US States. *Political Analysis*, 1–15.

Wilkerson, John, & Casas, Andreu. 2017. Large-scale computerized text analysis in political science: Opportunities and challenges. *Annual Review of Political Science*, **20**, 529–544.

Zhang, Ye, & Wallace, Byron C. 2015. A Sensitivity Analysis of (and Practioners' Guide to) Convolutional Neural Networks for Sentence Classification. *In: arXiv preprint arXiv:1510.03820.*