



A Quantitative Portrait of Legislative Change in Ukraine

Zachary K. Stine^(✉)  and Nitin Agarwal

University of Arkansas at Little Rock, Little Rock, AR 72204, USA
{zkstine, nagarwal}@ualr.edu

Abstract. Over the past decade, Ukraine has undergone tremendous socio-political changes, which continue to this day. While such changes may be analyzed and interpreted from a variety of sources, we utilize recent advancements in the quantitative analysis of culture to identify how these changes are encoded within Ukraine’s legislation. Our goal is to provide a new picture of Ukrainian governance that may be used by subject matter experts as a complement to existing forms of political data. To do so, we apply probabilistic topic modeling to compress over a decade of Ukrainian legislation into patterns of word usage. We then apply a recently developed calculation of novelty to measure how different each draft law is from the draft laws which precede it. We find an interesting pattern of legislative changes and identify some of the drivers of these changes. Finally, we discuss the relationship between our results and the broader context of Ukrainian political changes and suggest steps to explore this relationship further.

Keywords: Computational social science · Political science · Ukraine · Topic modeling · Legislation

1 Introduction

As a country, Ukraine comprises an ongoing series of socio-political changes. In particular, the parliament of Ukraine—the Verkhovna Rada—consists of an ever-changing array of political factions in which membership is fluid [1]. In this paper, we focus our analysis of Ukrainian politics between the years 2006 and 2018. Within this time period, multiple events with political salience occurred, including the 2014 ousting of Viktor Yanukovich as president. More broadly, this time period encompasses the fifth, sixth, seventh, and the majority of the eighth convocation, each representing the tenure of a newly-elected parliament.

In this study, we computationally analyze the draft legislation produced in each of the aforementioned convocations in order to examine how changing linguistic patterns within the legislation text might provide a complementary window into the country’s political changes during this time. This analysis is made possible by the public availability of documents relating to registered bills on the Verkhovna Rada website¹.

¹ <https://rada.gov.ua>.

To conduct our analysis, we apply topic modeling [2] to the corpus of draft laws in order to identify patterns of word usage contained within. We then calculate how novel each draft law is in light of the laws that preceded it, utilizing the measurement of novelty put forward by [3]. We find that periods of elevated average novelty exist which correspond to salient periods of political change within the country. We also show that convocations VI, VII, and VIII are each characterized by distinct trends in average novelty. We show that a series of draft laws related to how elections are conducted account for one distinct period of elevated average novelty, and we identify which parliamentary committees are most responsible for introducing novel legislation. These findings serve to paint a quantitative picture of legislative evolution based on language patterns, which emerge from a massive collection of documents. This resulting portrait serves as a useful complement to traditional political science analysis, providing a view of legislative change that is inaccessible through the close reading of a smaller number of documents.

An examination of legislative novelty is important for many reasons. A bill that is highly novel may represent the introduction of new legislative discourse or a new combination of extant legislative discourses. Such bills may provide early signals of legislative shifts and are therefore likely to be salient for a variety of analyses. Additionally, periods of higher and lower legislative novelty may indicate periods of legislative exploration in which the seeds of new legislative goals are planted or periods of focus on a particular legislative path respectively.

The remaining sections of this paper are organized as follows. In Sect. 2, we provide the necessary background for understanding our methodology and a brief review of other works which analyze political text. In Sect. 3, we describe the methodology used and present our results in Sect. 4. In Sect. 5, we discuss possible interpretations of these results and suggest further steps. A brief conclusion follows in Sect. 6.

2 Related Work

2.1 Topic Models and Political Text

To carry out our analysis, we use the topic modeling algorithm, latent Dirichlet allocation (LDA) [2], to identify a fixed number of word-usage patterns (*i.e.*, topics) in our corpus and represent each draft law as a distribution of these topics. Importantly, LDA can be thought of as operationalizing certain sociological concepts including framing, polysemy, heteroglossia, and a relational approach to meaning [4].

A broad overview of the use of computational methods for analyzing political text is provided by [5], which references examples of topic model extensions used to analyze speeches made in the U.S. Senate [6] and press releases from U.S. senators [7]. As illustrated in these examples, the use of topic models in analyzing political text often takes the resulting topics as the primary outputs for interpretation. In this study, however, the use of topic modeling is primarily a means to transform documents into low-dimensional, semantically useful representations in order to carry out additional calculations.

2.2 Textual Novelty

The methodology we employ centers on the notion of novelty put forward by [3]. In that paper, the authors applied LDA to a corpus of speeches made during the first parliament of the French Revolution. With each speech represented as a distribution of topics, they define and calculate two related measures: a speech’s novelty, N , and its transience, T . Both measures are based on the Kullback-Leibler divergence (KLD), which is an asymmetric measure of difference between two probability distributions also known as relative entropy. Novelty can be thought of as a quantity of how surprising a distribution is in light of the past, whereas transience relates to how surprising a distribution is in light of the future. In the present study, we only make use of novelty. The formulation of novelty given in [3] is itself based on a measure of textual novelty used within a cognitive framework describing how an information-seeking agent explores an environment of ideas [8]. A formal definition of novelty is provided in Sect. 3.

3 Methods

3.1 Data

As previously noted, a Verkhovna Rada website is maintained that enables users to view details about registered bills and download relevant documents, including the text of draft laws. For convocations V, VI, VII, and VIII, we download all available documents for each bill that corresponds to a draft law. In some cases, a single bill will have more than one available draft. In such cases, we download all available drafts. Curiously, there are a large number of draft laws from convocation V that do not have any draft documents available. For that reason, we restrict the bulk of our results to the convocations following it (*i.e.*, VI, VII, and VIII). However, the available draft laws from convocation V are still included in our analyses for two reasons: First, they provide useful training documents for constructing topic models, and second, they serve as a backdrop for calculating novelty for the early draft laws in convocation VI. The resulting corpus consists of 17,485 documents representing 17,164 draft law bills over a period of almost 12 years. A convocation-level breakdown of the corpus is given in Table 1.

Table 1. Description of data collected by convocation.

Convocation	Unique draft law bills	Total documents available	Earliest reg. date	Latest reg. date
V	514	533	May 25, 2006	Sep 4, 2007
VI	6,202	6,314	Nov 23, 2007	Dec 6, 2012
VII	3,888	4,015	Dec 12, 2012	Nov 26, 2014
VIII	6,560	6,623	Nov 27, 2014	May 17, 2018

In addition to these draft laws, we also make use of supplemental data generously provided to us by Dr. Tymofiy Mylovanov and his team at VoxUkraine². These supplemental data include voting results (positive or negative) and the main committee from which the draft law originated. We were able to cross-reference our analysis with 2,974, 701, and 1,144 draft laws from the supplemental data for convocations VI, VII, and VIII respectively.

3.2 Topic Modeling

After collecting the text of each available draft law, we then use LDA (as implemented in [9]) to infer some number of topics in order to represent each draft law as a distribution of topics. To do so, we perform some minimal preprocessing of the text. We tokenize the text and remove all punctuation. We remove a limited number of stopwords (*i.e.* functional words that carry little semantic information). This is done primarily for convenience when manually inspecting topics, given that stopword-removal ultimately has a superficial effect on topics [10]. We do not perform stemming, which has also been shown to have little effect on topic models and may even have negative effects [11].

In some cases, it may make sense to evaluate the number of topics chosen, k , in order to find an optimal value, where what constitutes ‘optimal’ is likely contingent on what question is motivating the use of topic modeling. Here, however, we are not necessarily concerned with precise interpretations of each topic, but rather with topic modeling as a useful way of reducing the dimensionality of the documents. We therefore explore multiple choices of k in order to assess how sensitive our results are to each choice. Here, we train topic models with 20, 40, 60, 80, and 100 topics.

3.3 Novelty

Once topic modeling has been performed, the novelty of a given draft law can be calculated following the methods in [3]: for some number of preceding documents, the novelty of a document’s topic distribution, d , is the average of the KLD from each preceding document’s topic distribution to d . In the case of draft laws, a particular law’s novelty represents how surprising that law’s topics are, given an expectation of the preceding laws’ topics. The number of preceding draft laws defines the scale at which novelty is computed and is denoted as the window width, w , a positive integer. The smallest possible scale, $w = 1$, is simply the KLD of draft law, $d^{(j)}$, relative to draft law $d^{(j-1)}$, which is given by

$$KLD\left(d^{(j)}|d^{(j-1)}\right) = \sum_{i=1}^K d_i^{(j)} \log_2\left(\frac{d_i^{(j)}}{d_i^{(j-1)}}\right). \quad (1)$$

² <https://voxukraine.org>.

For w greater than 1, the KLD is calculated between $d^{(j)}$ and each of the preceding w documents then averaged, so that the novelty of the j^{th} draft law, represented by its topic distribution, is given by

$$\mathcal{N}_w(j) = \frac{1}{w} \sum_{s=1}^w KLD(d^{(j)} | d^{(j-s)}). \quad (2)$$

Just as we explore various choices of k to understand how sensitive our results are to the number of topics, we similarly explore different choices of w to see the effects of the scale at which novelty is computed. In this study, we calculate novelty using $w = 50, 100, 200, 400, 800, 1,000,$ and $2,000$.

We order draft laws based on their registration dates, representing a bill’s formal birth into the legislative process. For days on which multiple bills were registered, we further order them based on their assigned bill number (however, this becomes more or less irrelevant for the scales at which we compute novelty).

4 Results

When comparing convocations VI, VII, and VIII in terms of novelty, we find that the mean novelty of convocation VI is greater than both VII and VIII. This finding is robust to changes in k and w (see Table 2).

Table 2. Comparison of novelty mean (std. deviation) for different values of k and w .

Convocation	$k = 20$		$k = 100$	
	$w = 100$	$w = 800$	$w = 100$	$w = 800$
VI	5.541 (2.333)	5.465 (2.322)	8.327 (2.373)	8.345 (2.357)
VII	4.916 (2.405)	4.872 (2.375)	7.525 (2.566)	7.433 (2.508)
VIII	4.997 (2.352)	4.999 (2.352)	7.613 (2.471)	7.628 (2.474)

In order to identify periods of especially high novelty over the ordering of bills, we plot each bill’s novelty along with a moving average line (Fig. 1). For bills that have multiple drafts, we only include the draft which has the highest novelty. This is because, given a law that has a draft with high novelty, each other draft is likely to also have high novelty. Including both drafts in this case would artificially inflate the average novelty.

When examining the moving average of novelty within a window of 100 laws, we see several interesting features. The bills registered at the beginning of both convocation VI and VII have elevated average novelty. This is especially true for the beginning of convocation VII. However, convocation VIII does not display elevated average novelty within its first few bills, but rather at the very end of 2017. This basic pattern—elevated average novelty at the beginnings of convocations VI and VII and in late 2017 of convocation VIII—is robust to changes in k and w (see Fig. 2).

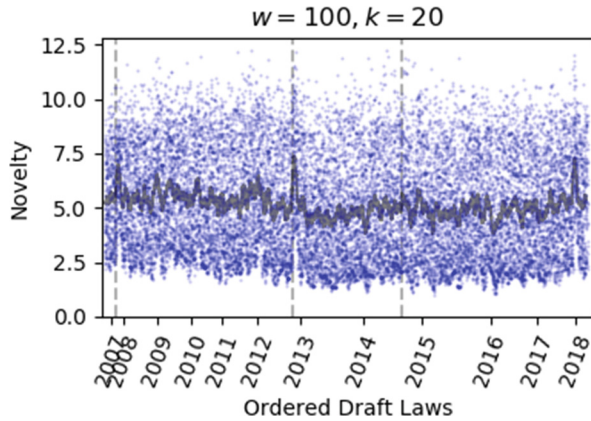


Fig. 1. A scatterplot of each draft law’s novelty calculated at the scale of $w = 100$ and with $k = 20$ along with a moving average line of novelty values from 100 draft laws. The vertical dotted lines, from left to right, represent the first bill registered in convocation VI, VII, and VIII.

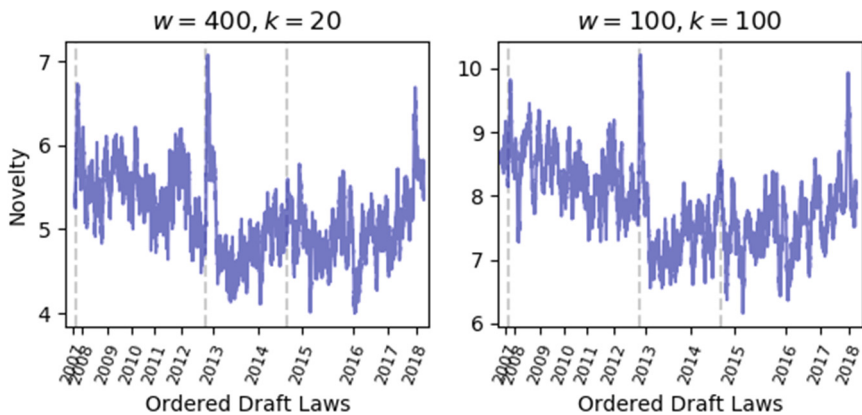


Fig. 2. Two examples of moving average lines within a window of 100 draft laws. Left: Moving average for novelty calculated at a scale of $w = 400$ and $k = 20$. Right: $w = 100$ and $k = 100$. While differences exist between the two, several features are stable across different values of w and k . Vertical dotted lines indicate the first bill of a new convocation.

In addition to these three peaks of average novelty, we also find that the average remains fairly elevated throughout much of convocation VI, which accords with our finding that convocation VI comprises higher novelty bills on average than convocations VII and VIII (see Table 2). What is notable when examining the moving average is that, after the peak in average novelty at the beginning of convocation VII, the average drops off significantly. After this drop, the average novelty oscillates between slight increases and more drops. Each drop in novelty signifies the presence of bills that exhibit similar topic distributions. This drop in average novelty is also robust to changes in k and w (Fig. 2).

While changes in legislative novelty are interesting in of themselves, it may also be of interest to examine precisely which bills and their corresponding topics most account for certain periods of interest. For example, we find that the period of elevated average novelty among bills registered in late 2017 is partly due to several draft laws on the subject of how elections are conducted (*e.g.*, bill 7366-1). Voters in Ukraine may cast their vote for a particular party without being provided a full list of candidates from the party, which has become the subject of a debate about whether such party lists should be made open³. These high-novelty draft laws are concerned precisely with this debate. Importantly, these laws were identified solely by their high novelty without prior knowledge of their political salience. While this may not necessarily be the case for every draft law with high-novelty, this example does suggest a link between legislative novelty and salience.

For the subset of bills for which voting and committee data was available, we make two comparisons. First, we compare the novelty of draft laws with positive voting results to those with negative voting results for both $k = 20$ and $k = 100$ with $w = 100$ in each case. We find that for both values of k , the mean novelty of draft laws with positive voting results is slightly less than the mean novelty of draft laws with negative results. However, this difference is much more pronounced for $k = 20$ than for $k = 100$ (Table 3).

Table 3. Comparison of novelty mean (std. deviation) for draft laws with voting results.

	Positive & negative results	Positive results	Negative results
$k = 20$	5.253 (2.447)	5.095 (2.407)	5.616 (2.500)
$k = 100$	8.4015 (2.510)	8.4014 (2.529)	8.4017 (2.468)

Second, we compare the novelty of draft laws based on which main committee produced the law. Here, we find that the choice of k is important. When comparing the ordering of each committee by its mean novelty for $k = 20$ and $k = 100$, we find fairly similar rankings of the committees except for two extreme cases: the two committees with the lowest mean novelty at $k = 20$ become the two most novel committees for $k = 100$. This is because when only 20 topics are available, the topics relevant for the types of bills produced by these two committees become subsumed by a broad and more generic topic. However, as the topic-granularity becomes finer with increasing values of k , topics more specific to these committees emerge as their own topics. Thus, we feel confident that the ordering based on $k = 100$ is much more informative than $k = 20$.

Excluding committees with fewer than ten bills available for analysis, the two committees with the highest average novelty are the Committee on European Integration (11.148 bits) and the Committee on Foreign Affairs (10.996 bits) (see Fig. 3). The Committee on Budget has the lowest mean novelty of 6.247 bits.

³ We thank Dr. Tymofiy Mylovanov for pointing this out to us.

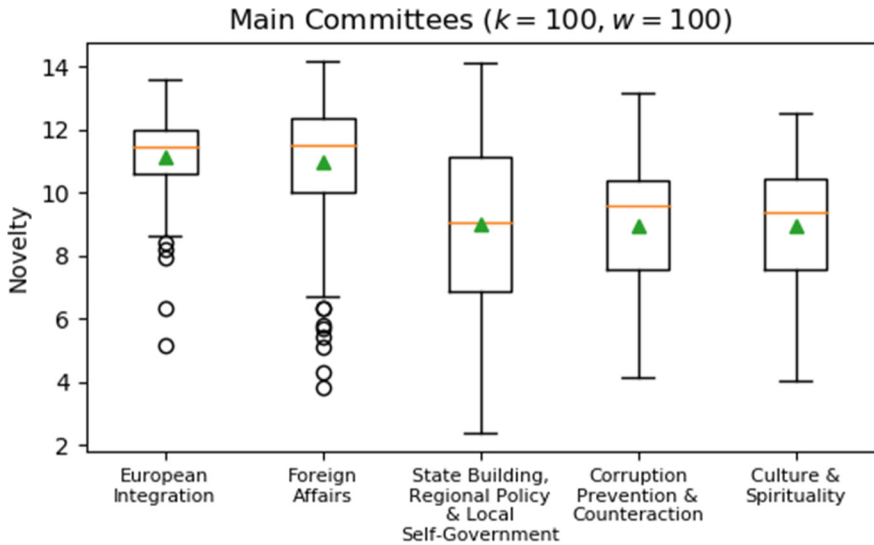


Fig. 3. A comparison of novelty distributions among the bills of the five main committees with the highest mean novelty values. Horizontal lines within the boxes denote the distribution median and triangles denote the mean. The committees are ordered from highest mean novelty on the left.

5 Discussion

Our findings suggest several interesting interpretations about the unfolding of the Ukrainian legislative process over the past decade. Each of the findings we have described illustrate how the analysis of legislative novelty may provide a complementary window into Ukrainian politics that is useful to political scientists. We therefore offer the following interpretations of our findings in the hope that they stimulate more in-depth analysis by political scientists and serve as an example of what types of interpretations become possible when analyzing legislative novelty.

First, we found that convocation VI comprises higher-novelty draft laws on average than convocations VII and VIII. This suggests that convocation VI dealt with a large number of newly encountered issues or dealt with extant issues in novel ways. Additionally, the elevated average novelty of convocation VI may signal the exploration of multiple legislative paths. The sharp increase in average novelty coinciding with the beginning of convocation VII may indicate a severe departure in legislative goals from convocation VI. This is supported by the fact that this peak in average novelty remains severe even at the scale of $w = 2,000$. The observed drop in average novelty directly after this peak indicates that, once the legislative direction changed at the outset of convocation VII, it stabilized. This is because low average novelty indicates repetition of the legislative topic distributions—the newly established legislative direction continues to be followed for a time. Interestingly, we see the average novelty begin a gradual ascent leading into the 2014 Ukrainian Revolution. Following this ascent, the oscillations in average novelty in convocation VIII indicate periods of

fixed legislative themes (low novelty) interspersed with periods of legislative change, culminating in the previously described peak in late 2017.

Second, we found that, among a subset of draft laws for which voting data was made available, there is a slight bias towards passing draft laws with lower-than-average novelty. Such a bias indicates a possible reluctance to pass bills that constitute severe departures from established legislative norms. However, this bias appears to diminish for larger values of k , so further analysis of voting data is needed.

Third, we found that, of the main committees in which draft laws originated, the Committee on European Integration and the Committee of Foreign Affairs are responsible for the highest novelty bills on average (again, for a subset of bills). This is notable in light of the 2013 protests which would lead up to the 2014 revolution and eventual ouster of President Viktor Yanukovich. The protests were initially motivated by the decision to break association talks with the European Union, widely seen as a capitulation to Russian interests⁴. The high average novelty of these committees suggests that they have been drivers of legislative innovation and change across these convocations. In other words, the greatest legislative changes undergone by Ukraine largely deal with how the country has managed its relationships abroad. While this interpretation may strike some as obvious, it is important to note that these committees were identified purely through quantitative means.

Several weaknesses exist in this current study, which we intend to address in future work. First, we have only considered bills representing draft laws. While draft laws constitute a great deal of all Ukrainian bills, the inclusion of other bill types will provide an even bigger picture of Ukraine's political evolution. Additionally, we will continue to increase the number of extra-textual features incorporated from each bill (e.g., bill sponsorship, supporting committees, etc.). Finally, we intend to more closely collaborate with relevant subject matter experts in order to bolster the interpretations of our results.

6 Conclusion

As the textual artifacts of a complex political process, Ukrainian draft laws encode the paths explored through a political space on the part of the Verkhovna Rada. By condensing each draft law into a distribution of inferred topics, we can measure how surprising a given law is relative to some number of preceding laws using the notion of novelty from [3]. When we analyze approximately twelve years of draft laws in this way, an interesting picture of Ukrainian political evolution emerges: a period of high average novelty throughout convocation VI, a sharp increase in average novelty followed by low average novelty throughout convocation VII, and a gradual increase in average novelty throughout convocation VIII reaching a crescendo in late 2017. We also see that, of the parliamentary committees, those that are most responsible for driving legislative changes are those that deal with how Ukraine relates to other countries.

⁴ For an example of how this was reported in American media, see <https://www.npr.org/sections/thetwo-way/2013/11/25/247184300/ukraine-protests-continue-over-suspension-of-eu-talks>.

Acknowledgements. The authors wish to thank Dr. Tymofiy Mylovanov for explaining the political context of certain bills and for the supplemental draft law data provided by him and his team at VoxUkraine. This research is funded in part by the U.S. Office of Naval Research (N00014-17-1-2675) and the Jerry L. Maulden-Entergy Endowment at University of Arkansas – Little Rock. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding organizations. The authors gratefully acknowledge the support.

References

1. Herron, E.S.: Causes and consequences of fluid faction membership in Ukraine. *Europe-Asia Stud.* **54**(4), 625–639 (2002)
2. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent Dirichlet allocation. *J. Mach. Learn. Res.* **3**, 993–1022 (2003)
3. Barron, A.T.J., Huang, J., Spang, R.L., DeDeo, S.: Individuals, institutions, and innovation in the debates of the French Revolution. *PNAS* **115**(18), 4607–4612 (2018)
4. DiMaggio, P., Nag, M., Blei, D.: Exploiting affinities between topic modeling and the sociological perspective on culture: application to newspaper coverage of U.S. government arts funding. *Poetics* **41**, 570–606 (2013)
5. Grimmer, J., Stewart, B.M.: Text as data: the promise and pitfalls of automatic content analysis methods for political texts. *Polit. Anal.* **21**(3), 267–297 (2013)
6. Quinn, K.M., et al.: How to analyze political attention with minimal assumptions and costs. *Am. J. Polit. Sci.* **54**(1), 209–228 (2010)
7. Grimmer, J.: A Bayesian hierarchical topic model for political text: measuring expressed agendas in Senate press releases. *Polit. Anal.* **18**(1), 1–35 (2010)
8. Murdock, J., Allen, C., DeDeo, S.: Exploration and exploitation of Victorian science in Darwin’s reading notebooks. *Cognition* **159**, 117–126 (2017)
9. Řehůřek, R., Sojka, P.: Software framework for topic modelling with large corpora. In: *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, pp. 45–50 (2010)
10. Schofield, A., Magnusson, M., Mimno, D.: Pulling out the stops: rethinking stopword removal for topic models. In: *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: vol. 2, Short Papers*, pp. 432–436. ACL, Spain (2017)
11. Schofield, A., Mimno, D.: Comparing apples to apple: The effects of stemmers on topic models. *Trans. Assoc. Comput. Linguist.* **4**, 287–300 (2016)