

Topological Data Analysis for Discourse Semantics?

Ketki Savle
UNC Charlotte
kropleka@uncc.edu

Wlodek Zadrozny
UNC Charlotte
wzadroz@uncc.edu

Minwoo Lee
UNC Charlotte
Minwoo.Lee@uncc.edu

Abstract

In this paper we present new results on applying topological data analysis (TDA) to discourse structures. We show that topological information, extracted from the relationships between sentences, can be used in inference, namely it can be applied to the very difficult legal entailment problem given in the COLIEE 2018 data set. Previous results of Doshi and Zadrozny (2018) and Gholizadeh et al. (2018) show that topological features are useful for classification. The applications of computational topology to entailment are novel, and in our view provide a new set of tools for discourse semantics: computational topology can perhaps provide a bridge between the brittleness of logic and the regression of neural networks. We discuss the advantages and disadvantages of using topological information, and some open problems such as explainability of the classifier decisions.¹

1 Introduction

Topology is a classic branch of mathematics that deals with shape invariants such as the presence and numbers of holes. More recently *topological data analysis* (TDA) was introduced as a branch of computational mathematics and data science, predicated on the observation that data points have implicit shapes (e.g. Edelsbrunner and Harer (2010)). Throughout the paper we will be using the word *topology* only in these two particular senses.

Both topology and TDA can be viewed as an abstraction mechanism, where we replace the original shape or cloud of data points by some numbers representing their mathematical properties, using a formal machinery derived from algebraic topology. In case of TDA, we use software implementing these methods.

A natural question to ask is whether texts or discourse structures have shapes that can be measured using tools of topology. Zhu (2013) was the first to investigate this question and observed we can capture some information about discourse structures using topological structures, namely homological persistence (which we do not have space to define here, and we simply use it as a source of numerical features). Zhu used a collection of nursery rhymes to illustrate how topology can be used to find certain patterns of repetition. More recently, Doshi and Zadrozny (2018) applied Zhu’s method in a larger setting showing its classification superiority on the task of assigning movie genres to user generated plot summaries, using the IMDB data set. They improved on the early 2018 state of the art results of Hoang (2018), which was achieved using deep learning on this large data set. Gholizadeh et al. (2018) applied a different method for computing homological persistence to the task of authorship attribution, which is also a classification task, showing that the patterns of how authors introduce characters in novels can be captured to large extent using topological descriptors. Interestingly, neither of these works uses topological features to augment the usual *tf/idf* representations of documents: Doshi and Zadrozny (2018) use counts of words (from a previously identified vocabularies) to form a matrix which is the only input to topological persistence, and then they make a rule based decision based only on the presence of barcodes; and Gholizadeh et al. (2018) use time series. To use topological data analysis (TDA), Zhu (2013) assumes

¹This paper has been accepted for publication and presentation at the *13th International Conference on Computational Semantics (IWCS 2019)*; University of Gothenburg, Sweden; 23-27th May 2019. IWCS is the bi-yearly meeting of SIGSEM, the ACL special interest group on semantics.

that text is implicitly coherent (SIFTS method), and so do Doshi and Zadrozny (2018). Namely, they assume implicit connection between consecutive sentences in each document. While for movie plots this assumption makes sense, it might be more problematic in other contexts, such as entailment, especially when two passages are unrelated.

1.1 Our results

In this paper, we present our very recent results on applying topological data analysis (TDA) to entailment, with some improvement of accuracy over the baseline without persistence.

More specifically, this paper shows TDA works on entailment improving the task of classification for establishing entailment on the COLIEE 2018 task by over 5% (F-measure) compared to the results classification without topology that is using only tf/idf and similarity. Furthermore, this result does not assume the existence of the implicit skeleton connecting consecutive sentences (as was done in Doshi and Zadrozny (2018), following Zhu (2013)).

The title of the present article ends with a question mark. This question mark reflects the tension between the positive empirical results derived using topological methods and our lack of understanding why these methods work. Thus, perhaps another contribution of this paper is to point to both, the need for theoretical inquiry about relationships between discourse and its topological abstractions, and more importantly to the need for tools that would allow us to experiment with such hypothetical relations. As we speculate in Section 4, the effectiveness of TDA for entailment *might* be explainable using the known mathematical connections of topology and logic (e.g. Vickers (1996)). Proper tooling could prove or disprove this hypothesis.

1.2 A minimum background on topological data analysis

Topological Data Analysis (TDA) can be viewed as a method of data analysis done at different resolutions. Informally speaking, this process can be viewed as data compression(cf. Lum et al. (2013)). It can also be viewed as an attempt to reconstruct shape invariants, such as presence of voids or holes, from collection of points, at different resolutions (Edelsbrunner et al. (2000)). Or in yet another formulation TDA tries to make data points fit together, and measures their divergence from perfect fit Robinson (2014) (we will not be using this last property here).

Figure 1 (taken from Huang et al. (2018)) conveys these ideas: it shows a cloud of data points, and its subsequent approximation by balls of increased radii. The overlaps produce a change in shape which can be measured using the H_0 and H_1 lines: The number of H_0 lines intersecting the vertical bar at ϵ is the number of connected components of when the points are extended with balls of that radius. Therefore as ϵ increases, the number of components decreases. In this process the exact values of the data points are ignored, but the shape information is preserved – that is, two clouds of similar shapes but different values will have similar persistence diagrams. The H_1 lines show the birth and death of holes at given values of ϵ . The top line show a hole persisting from 1.2 to 3.3 (approximately). Jointly, H_0 and H_1 (and higher H_n 's, not discussed here) compress information about the shape of the point cloud. This diagram deals only with planar structures, but persistence works in higher dimensions as well, in principle allowing machines to "see" shapes in dimensions higher than 3, a task difficult for humans. "Persistence" refers to the fact that the number of components and holes remains stable at some intervals, and we record this fact as numerical features; "homology" means similarity (of shape).

In NLP, the points are in a high dimensional space and represent vectors of tf/idf or other features derived from text. The method works the same, but *please note that Figure 1 only illustrates how TDA progresses from points to shapes. At this point, we do not know — and we see it as a major open problem — what aspects of natural language semantics, whether for entailment or classification, are captured by topological features.* (Although, as mentioned earlier, some aspects of this problem are discussed in Zhu (2013)).

To finish this introduction, we mention an equivalent representation, called *persistence diagram*, an example of which appears later in Figure 5, represents birth and death as two dimensional coordinates,

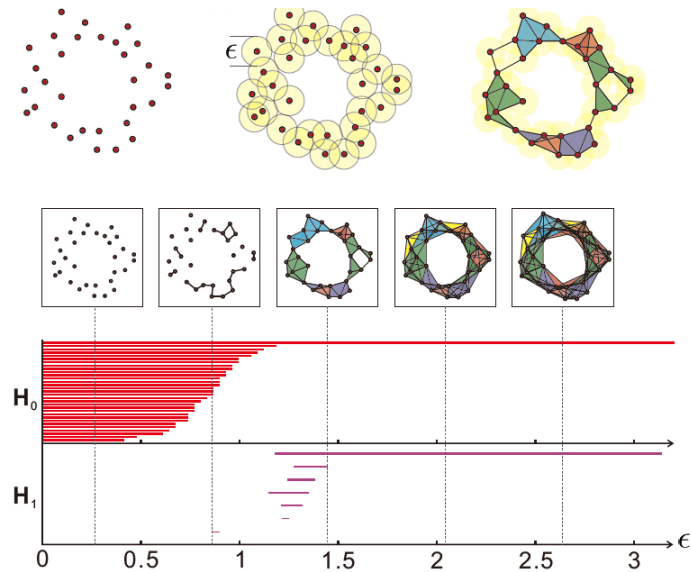


Figure 1: Persistence homology is a data analysis tool. Intuitively, as we start expanding the data points into balls of increased radii, planar figures emerge and change. The intervals in H_0 and H_1 capture relevant features of this process, namely the number of connected components, and the number of holes at different resolutions. The method abstracts distance information about the feature vectors of original data. It is an open problem how exactly these new numerical features help entailment. Source of the figure: Huang et al. (2018)

and uses colors to make a distinction between H_0 and H_1 . To repeat, the representation method is general, and it generates numbers we can use as machine learning features. However, finding the corresponding natural language mechanisms responsible for the improvements in accuracy of classification or entailment is an open problem.

1.3 Related work on applying topological data analysis to discourse modeling, and text processing in general

Applications of TDA to text started with discourse: Zhu (2013) used nursery rhymes to illustrate properties of homological persistence (e.g. that it is not simply measuring repetitions), and also showed that children, adolescent and adult writing styles can be differentiated using TDA. Doshi and Zadrozny (2018) used Zhu’s tools and methods to show that topological features can improve the accuracy of classification (movie plots). They also discuss the paucity of applications of TDA to text, and the fact that not all of these applications show improvements over the state of the art: in particular this was the case for sentiment analysis and clustering Michel et al. (2017). Temčinas (2018) argues for applicability of persistent homology to lexical analysis using word embeddings, and in particular for discovery of homonyms such as ‘bank’, thus potentially for word sense disambiguation.

For discourse analysis, broadly speaking, we see that according to Guan et al. (2016) TDA can help with extraction of multiword expressions and in summarization; also it might be worth to mention Horak et al. (2009) apply TDA to a networks of emails, but without going into their text. In other words, TDA for text data is an emerging area of research, perhaps with a potential to be of value for computational linguistics (see the last two sections of this paper for an additional discussion).

2 Entailment between legal documents

The COLIEE task: Our application of topological data analysis (TDA) to computing entailment focuses on the legal entailment COLIEE ² task, i.e. *Competition of Legal Information Extraction and*

²COLIEE 2018 Workshop collocated with JURISIN 2018: <https://sites.ualberta.ca/~miyoung2/COLIEE2018/>

Entailment (COLIEE).

To solve an entailment task, given a decision of a *base case*, along with its summary and facts, the system should be able to establish the relation of entailment with an associated *noticed case*, given as a list of paragraphs. We can define it as, given a base case b , and its decision d , and another case r represented by its paragraphs $P = \{p_1, p_2, p_3, \dots, p_n\}$, and we need to find the set $E = \{p_1, p_2, \dots, p_m \mid p_i \in P\}$, where $entails(p_i, d)$ denotes a relationship which is true when $p_i \in P$ entails the decision d (c.f. Rabelo et al. (2018), Kim et al. (2016), Adebayo et al. (2016)).

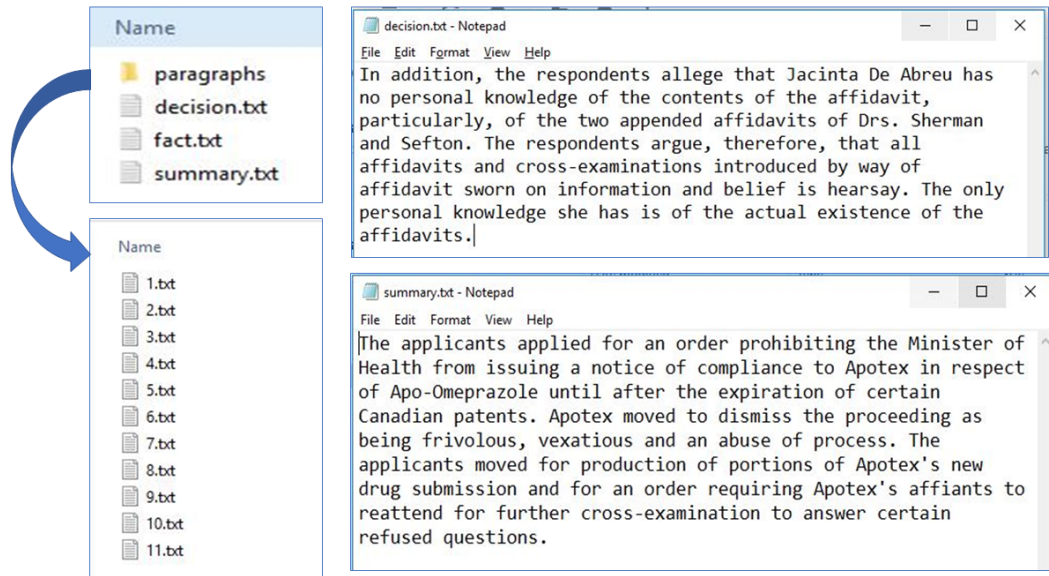


Figure 2: Each case folder includes decision file, summary file and fact file with paragraph folders. Decision file is an actual query i.e. a decision of a base case, summary file consists of a summary of a base case and facts file includes all the human annotated facts about the base case.

Text	Hypothesis
<p>As previously stated, the applicant contends that the CRDD based its decision that his testimony lacked credibility or trustworthiness by drawing inferences unsupported by the evidence and by ignoring portions of the applicant's testimony and documentary evidence A Convention refugee claimant applied for judicial review of the dismissal of his claim by the Convention Refugee Determination Division of the Immigration and Refugee Board.</p>	<p>It is my opinion that the Board acted arbitrarily in choosing without valid reasons, to doubt the applicant's credibility concerning the sworn statements made by him and referred to supra. When an applicant swears to the truth of certain allegations, this creates a presumption that those allegations are true unless there be reason to doubt their truthfulness. See: Villaroel v. M.E.I. (1979), 31 N.R. 50, and more particularly footnote number 6 to the Reasons of Pratte, J. On this record, I am unable to discover valid reasons for the Board doubting the truth of the applicant's allegations above referred to.</p>

Figure 3: Above example illustrates entailment between text and hypothesis for one of the base cases of COLIEE 2018. Text column consists of decision and summary of a base case and the hypothesis is an entailed supporting paragraph for a given base case. (We have excluded facts file in text while demonstrating as size of its text is large)

Overview of dataset: For training, there were 181 base cases provided which were drawn from an existing collection of Federal Court of Canada law cases. Every case consists of a decision file, summary file, facts file and a list of paragraph files. The training data also consists of labels in XML format for

entailed paragraphs. Our task was to identify paragraphs from this list, that entails with the decision of a base case. In 181 base cases, the number of paragraph files were 8794 out of which 239 were positively entailed and the rest were not entailed. This led us to a very imbalanced class ratio of 2.71% examples in positive class and 97.29 % in negative class.

Why this task is difficult: Since the data is of legal domain, it might require an understanding of law to analyze it: A traditional approach such as training neural network, or the more intuitive semantic similarity approach did not work very well on this dataset. Reason being, pre-trained word embedding such as GloVe and word2vec may not contain enough legal terms for neural networks to learn. Similarity correlates with entailment, but it clearly is a different problem. Also, this corpus is too small to use it to create our own pre-trained word embeddings. And at this point we do not have the bandwidth to pursue corpus expansion and create appropriate legal embeddings. An example of the type of text present in the COLIEE data is shown in Fig.3.

Another challenge was data distribution. Using common re-sampling techniques for classification task along with tf/idf leads to predicting always the negative class and treating positive class as noise, giving false high accuracy.

The best results obtained on COLIEE leaderboard was of Rabelo et al. (2018) where they employed similarity-based feature vector and used a “candidate” paragraph, chosen from histogram of the similarities between each noticed case and all paragraphs for classification. In this method, due to the unstructured input format, their team used post processing for classifier’s predictions. In case of too many positive detections, they retained 5 candidate paragraphs whereas for zero positive predictions they retained 1 paragraph by choosing classifier’s confidence interval. With this approach they delivered 0.24 precision, 0.28 recall and 0.26 F-score.

3 Computing entailment with and without topological features

To see whether topological features provide any additional information we employed a supervised machine learning approach. We represented the data points as a set of elements of type “[text, hypothesis], Label“. We defined ”text” as a combination of decision file, summary file, and fact file; and ”hypothesis” as a list of paragraphs for a case. For cleaning the text data, we simply removed punctuation, stop-words followed by converting the text to lower case and stemming it. This process, together and the features used in the experiments are shown in Fig. 4.

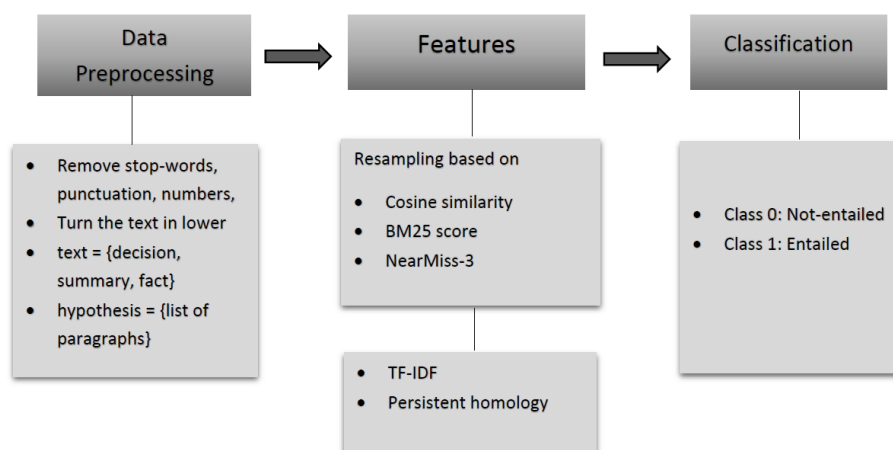


Figure 4: Diagram represents pipeline used for establishing entailment. A simple flow was to pre-process the data, prune highly similar and relevant paragraphs and resample further using NearMiss-3, then in the second pass, use homology features along with tf/idf.

We then formulated the problem as a binary classification problem for establishing corresponding paragraphs as entailed or not entailed with a base case. Mathematically, given a training data $D = (x_i, y_i)$

for $i = 1, \dots, N$, where $x_i = \{texts, hypothesis\}$, and $y_i = \{0, 1\}$.

3.1 Method 1: Relevance and similarity approach.

Considering every case had a list of paragraphs and severe imbalance, we approached this problem by first ranking the paragraph files using Okapi BM25 algorithm. We also calculated cosine similarity of a text and its hypothesis and combined these features to re-sample the data which we hoped will maximize the probability of establishing entailment without any information loss. This experiment was performed using Random Forest classifier for binary classification. The results, shown in Table 1 show improvement over previously reported top score of Rabelo et al. (2018) – note, however, our results were obtained after the JURISIN 2018 competition. Our main point is to see whether topological features provide additional value.

3.2 Method 2: Topological Data Analysis approach.

We discovered an opportunity to examine if topology can explain any structure which could create stronger signals to capture entailment. From the previous method we learned that entailment cannot be explained by establishing similarity only. By measuring the distance between two documents, one cannot necessarily infer a meaning of one text from another. In Information Retrieval, if a document is relevant to a given query, it does not necessarily mean that the meaning of a query can be completely inferred from the retrieved document. In fact, this creates a need for entailment in various NLP tasks including IR.

We used *Ripser*, a C++ library to compute persistent homology, for establishing topological structure of documents.³ Ripser was applied both to text and hypothesis. Our assumption is if text is entailed with hypothesis then the corresponding values of birth, death radius could provide stronger signals to the classifier. Unlike the movie classification experiment, we did not observe any specific barcode structure for entailed and non-entailed paragraphs, but the radius of birth-death cycle is significantly different for entailed documents as compared to the non-entailed ones. Another reason for not having a specific structure between such documents could be the length of these documents as each file consists of 5 sentences on an average. In future we aim to perform this experiment on larger size documents to see if there is any obvious barcode structure between entailed documents that can visually give us a clear interpretation. After calculating homology, we combined persistent homology features with tf/idf to create a feature vector comprised of the same. We used Random Forest classifier for binary classification task to establish entailment. Notably, we have not assumed the existence of coherence skeletons in documents (SIFTS in Zhu (2013)).

Experiment and Results:

We used tenfold cross-validation, setting a random sample of 22 cases aside from given 181 cases for the evaluation task. From our first method where we used highly similar and relevant paragraphs for classification along with tf/idf feature vector, our best results were 0.28 precision score, 0.58 recall and 0.38 F-score for entailed class (see Table 1).⁴ We improved our precision score by 2.5%, recall by over 14% and F-score by over 5% using topological data analysis. (Our aim was to achieve higher F-score for classification other than recall as a naïve implementation can give 1.0 recall by predicting all paragraphs as entailed). Using topological features, we could see reduction in predicting false positives, and more accurate predictions for true positives. We experimented with three machine learning classifiers out of which we obtained the best results using Random Forest.

³<https://github.com/Ripser/ripser>

⁴These results were obtained after the COLIEE competition.

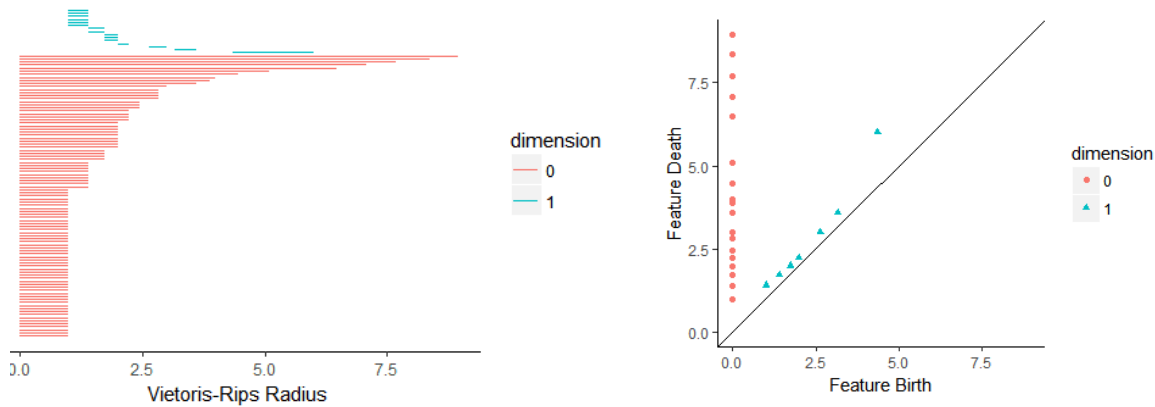


Figure 5: Left Panel: Barcode structure of persistent homology capturing multiple cycles. Note the the differences in radius of one long cycle and the others. Right panel: Persistent diagram representing the cycles from the left panel. Note the dot further from the diagonal corresponding to the long cycle. We show that these cycles are informative (Table 1), but we do not have tools to understand precisely how.

Method	Precision	Recall	F-score
Robelo et al. (2018) [prior art]	24%	28%	26%
Similarity + relevance score + tf/idf+ RF	28.2	58.3	37.6
Similarity + relevance score + tf/idf + RF + Topology	30.7	72.5	43.0

Table 1: Results of the classification experiments using Random Forest (RF) with 10-fold validation; RF produced best results, with and without topological features. In the first experiment, using proper filtering and resampling improved the F-score compared with COLIEE 2018 prior art. More importantly, we see that *the presence of topological features is informative for entailment* – this is the main point of the paper.

4 Discussion and Open Problems

As shown in Table 1, the use of topological features, namely birth-death information shown in Fig. 5, can improve the accuracy of computing entailment. However, it is an open issue to understand what exactly is being captured by using persistence. This can be seen as two sets of open problems: (a) we do not know exactly the correspondence between text and homological features; (b) we do not have instruments to capture these relationships.

We understand these relationship on some the abstract, mathematical level, even for text; in Zhu (2013) and Doshi and Zadrozny (2018) experiments, because of the simple setups, the 1-dimensional persistence measures the tie backs of content words. However, this is less clear for entailment, and we do not have instruments that would allow us to go back from the classifier decision and show the meaning of the topological features in documents we were using. *Thus the abstract and concrete explainability of topological text features is an open problem.* In addition, as the referees observed, entailment has direction, but distances used by our out of the box TDA methods are symmetric. So, what *exactly* is happening? – We don’t know. However, asymmetric structures as in Fig. 1 can arise from (symmetric) distances between points. One hypothesis we plan to explore is that ”global alignment” of Dagan et al. (2010) is captured by homological persistence. Similarly, it is conceivable that feature inclusion measures such as APinc, balAPinc, see e.g. Baroni et al. (2012), are indirectly captured by homological persistence. Again, it is an open problem what exactly is happening here.

In principle asymmetric measures of distance can be used in computational topology, see: Bubenik and Vergili (2018) and also discussion in Hennig and Liao (2013). Whether doing so would help entailment is an open problem.

To continue with speculations, there is a category theoretical style of research on entailment and distributional semantics, e.g. Bankova et al. (2016). There are also deep connections between topology, category theory and logic (e.g. Vickers (1996)). And we could even add physics to the mix: Baez and

Stay (2010). Given the connections between intuitionistic logic, Heyting algebras and topology, and the possibility of translation between these three representations (Vickers (1996)), we can speculate if we properly do computational topology for inference, we should get approximately-correct intuitionistic, logical inference methods. This could be an important connection, since logics are proverbially brittle, and computational topology is not. Thus our results *might* be experimentally confirming this intuition, and on a pretty difficult data set.

5 Summary and Conclusion

TDA can be computationally expensive, as observed by many researchers, and also Huang et al. (2018) to argue that quantum computing methods might be appropriate (if they materialize). However topological features seem to provide advantage when only small amount of the data is available, as shown here, and also in Doshi and Zadrozny (2018), who used only small percentage of data for preparation and training. This is also the case in our related work (Savle and Zadrozny (2019)), where we improved on Doshi and Zadrozny (2018) movie plot classification results by changing the inputs to the computation of persistent homologies from binary matrices to tf/idf representations augmented with persistence, which is the representation used here. Furthermore, we did not use the assumption of time skeleton Zhu (2013). From discourse interpretation point of view, this shows the assumption of discourse coherence does not have to be built in into the TDA method. But, again, the trade-offs between these two approaches are unclear.

Similarly, if larger amounts of data are given (e.g. movie plots), the precise computational tradoffs between using topology versus deep neural networks are unclear, especially given the ongoing improvements on various text analysis benchmarks, and new methods for addressing these tasks appearing on a daily basis.

Our future work includes, in the near horizon, experimenting with other data sets, possibly using graph embeddings in addition to topology. In a slightly longer horizon, we also want to explore higher dimensional persistence, which was shown in Horak et al. (2009) to capture relevant properties of a social network (email exchanges), but has not, to our knowledge, been used for other aspects of discourse understanding. And in parallel, we will be focusing on building tools to help us answer the question what exactly is captured by topological features.

In summary, this work confirms the ability of topological features to effectively capture certain structural properties of discourse text. On the one hand, it is another application of topological data analysis to text. On the other hand, given the paucity of positive results in this space (as discussed in the Introduction), and no previously reported applications to inference, we see our work as giving a new tool for computational discourse semantics, which could be used, as we have shown, as an addition to existing tools. Therefore, in our view, this research opens a new area of discourse analysis, where regression-based tools (such as standard machine learning and neural networks) can be used jointly with structural tools: to logic and ontology we can therefore add topology. From a formal point of view, with the known correspondence between intuitionistic logic and topology, the effectiveness of computational topology for inference, should yield approximate (and mostly correct) inference methods. This work shows that indeed this might possible, even for relatively difficult cases of entailment.

Acknowledgments: We thank the referees of IWCS 2019 for their comments and suggested improvements. Most of the issue raised by them we addressed in the preceding section. Unfortunately, explaining why exactly topological methods work on entailment is an open problem.

Authors' contributions: K. Savle designed, ran and analyzed the results of the experiments under the guidance of W. Zadrozny, and with additional help from M. Lee, esp. in analyzing machine learning results. KS and WZ were the primary writers of this paper.

References

- Adebayo, K. J., L. Di Caro, G. Boella, and C. Bartolini (2016). TEAMNORMAS's participation at the coliee 2016 bar legal exam competition, (submission id: N01). In *Tenth International Workshop on Juris-informatics (JURISIN)*.
- Baez, J. and M. Stay (2010). Physics, topology, logic and computation: a rosetta stone. In *New structures for physics*, pp. 95–172. Springer.
- Bankova, D., B. Coecke, M. Lewis, and D. Marsden (2016). Graded entailment for compositional distributional semantics. *arXiv preprint arXiv:1601.04908*.
- Baroni, M., R. Bernardi, N.-Q. Do, and C.-c. Shan (2012). Entailment above the word level in distributional semantics. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*, pp. 23–32. Association for Computational Linguistics.
- Bubenik, P. and T. Vergili (2018). Topological spaces of persistence modules and their properties. *Journal of Applied and Computational Topology*, 1–37.
- Dagan, I., B. Dolan, B. Magnini, and D. Roth (2010). Recognizing textual entailment: Rational, evaluation and approaches—erratum. *Natural Language Engineering* 16(1), 105–105.
- Doshi, P. and W. Zadrozny (2018). Movie genre detection using topological data analysis and simple discourse features. In *Proc. 6th International Conference on Statistical Language and Speech Processing, SLSP 2018, Vol. 11171 of Lecture Notes in Computer Science, Springer*.
- Edelsbrunner, H. and J. Harer (2010). *Computational topology: an introduction*. American Mathematical Soc.
- Edelsbrunner, H., D. Letscher, and A. Zomorodian (2000). Topological persistence and simplification. In *Foundations of Computer Science, 2000. Proceedings. 41st Annual Symposium on*, pp. 454–463. IEEE.
- Gholizadeh, S., A. Seyeditabari, and W. Zadrozny (2018). Topological signature of 19th century novelists: Persistent homology in text mining. *Big Data and Cognitive Computing* 2(4), 33.
- Guan, H., W. Tang, H. Krim, J. Keiser, A. Rindos, and R. Sazdanovic (2016). A topological collapse for document summarization. In *Signal Processing Advances in Wireless Communications (SPAWC), 2016 IEEE 17th International Workshop on*, pp. 1–5. IEEE.
- Hennig, C. and T. F. Liao (2013). How to find an appropriate clustering for mixed-type variables with application to socio-economic stratification. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 62(3), 309–369.
- Hoang, Q. (2018). Predicting movie genres based on plot summaries. *arXiv preprint arXiv:1801.04813*.
- Horak, D., S. Maletić, and M. Rajković (2009). Persistent homology of complex networks. *Journal of Statistical Mechanics: Theory and Experiment* 2009(03), P03034.
- Huang, H.-L., X.-L. Wang, P. P. Rohde, Y.-H. Luo, Y.-W. Zhao, C. Liu, L. Li, N.-L. Liu, C.-Y. Lu, and J.-W. Pan (2018). Demonstration of topological data analysis on a quantum processor. *Optica* 5(2), 193–198.
- Kim, M.-Y., R. Goebel, Y. Kano, and K. Satoh (2016). Coliee-2016: evaluation of the competition on legal information extraction and entailment. In *International Workshop on Juris-informatics (JURISIN 2016)*.

- Lum, P. Y., G. Singh, A. Lehman, T. Ishkanov, M. Vejdemo-Johansson, M. Alagappan, J. Carlsson, and G. Carlsson (2013). Extracting insights from the shape of complex data using topology. *Scientific reports* 3, sre01236.
- Michel, P., A. Ravichander, and S. Rijhwani (2017). Does the geometry of word embeddings help document classification? a case study on persistent homology based representations. *arXiv preprint arXiv:1705.10900*.
- Rabelo, J., M.-Y. Kim, H. Babikar, R. Goebel, and N. Farruque (2018). Information extraction and entailment for statute law and case law. In *JURISIN 2018*. Available at <http://research.nii.ac.jp/jurisin2018/>.
- Robinson, M. (2014). *Topological signal processing*. Springer.
- Savle, K. and W. Zadrozny (2019). Topological data analysis for text classification. *under review*.
- Temčinas, T. (2018). Local homology of word embeddings. *arXiv preprint arXiv:1810.10136*.
- Vickers, S. (1996). *Topology via logic*, Volume 5. Cambridge University Press.
- Zhu, X. (2013). Persistent homology: An introduction and a new text representation for natural language processing. In *IJCAI*, pp. 1953–1959.