CIRCLE 2022

Exploring SBERT and Mixup Data Augmentation in Rhetorical Role Labeling of Indian Legal Sentences

Alexandre G. de Lima
Mohand Boughanem
Eduardo Henrique da S. Aranha
Taoufiq Dkaki
José G. Moreno

INSTITUTO FEDERAL
Rio Grande do Norte

UFRN
UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE

IRIT
Institut de Recherche
en Informatique de Toulouse
CNRS - INP - UT3 - UT1 - UT2J

UNIVERSITÉ TOULOUSE III
PAUL SABATIER
Université de Toulouse

UNIVERSITÉ TOULOUSE
Jean Jaurès

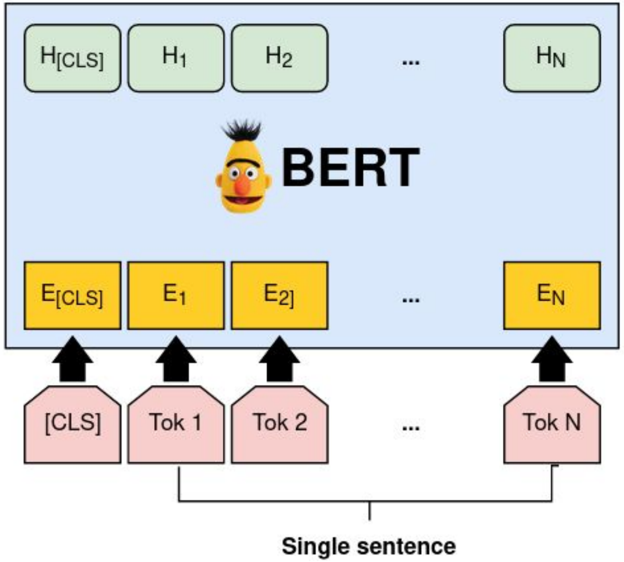
Introduction

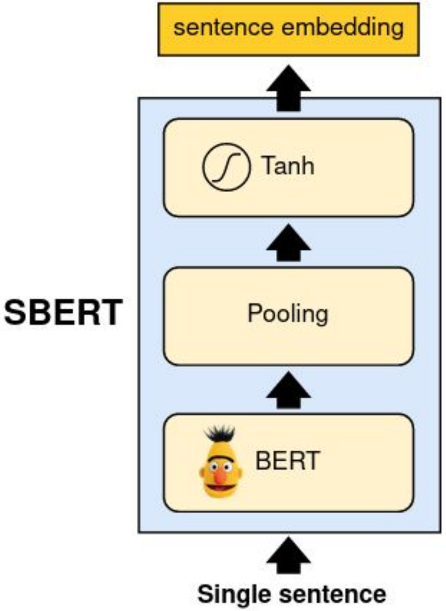
- The semantic recognition of text pieces from legal documents is a valuable task
- Our task consists of assigning rhetorical roles (**sentence classification**) to text sentences
- Pre-trained deep learning models like BERT and GPT has boosted general-domain NLP applications but legal applications can benefit from them too

Sentence	**Role**
The University had granted conditional approval, as stated earlier	Argument
S.11 of the Act deals with inspection	Statute
At that time in 1984,the AICTE Act was not on the statute book	Precedent

State of the Art

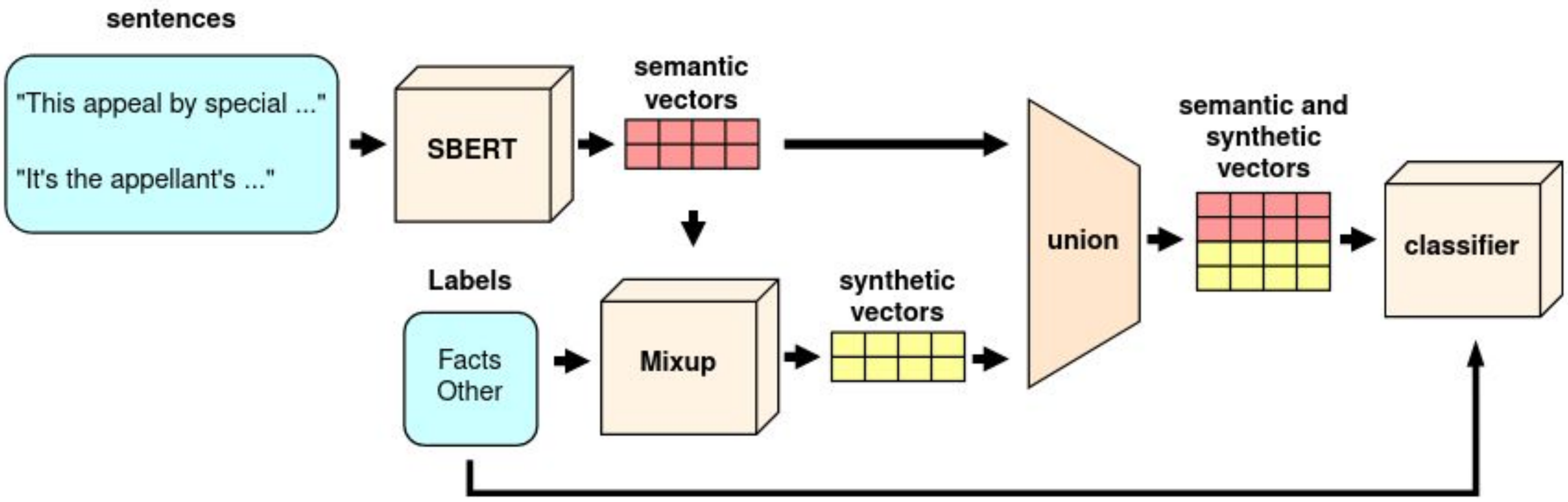
- Sentence BERT (SBERT): a model capable to generate semantically meaningful sentence embeddings from BERT hidden states in a computationally efficient way (Reimers and Gurevych, 2019).





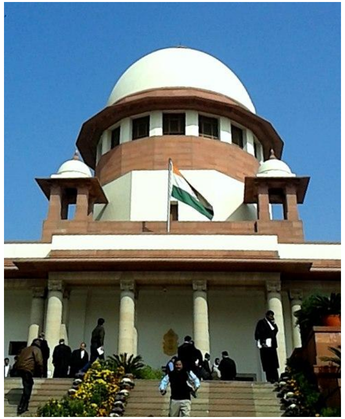
Proposed solution: exploit SBERT and MixUp

- Mixup: a data augmentation method agnostic to the nature of data or application. It also works as a regularization method (Zhang et al, 2018).


sentences
"This appeal by special ..."
"It's the appellant's ..."
SBERT
semantic vectors
Labels
Facts
Other
Mixup
synthetic vectors
union
semantic and synthetic vectors
classifier

Experimental setup

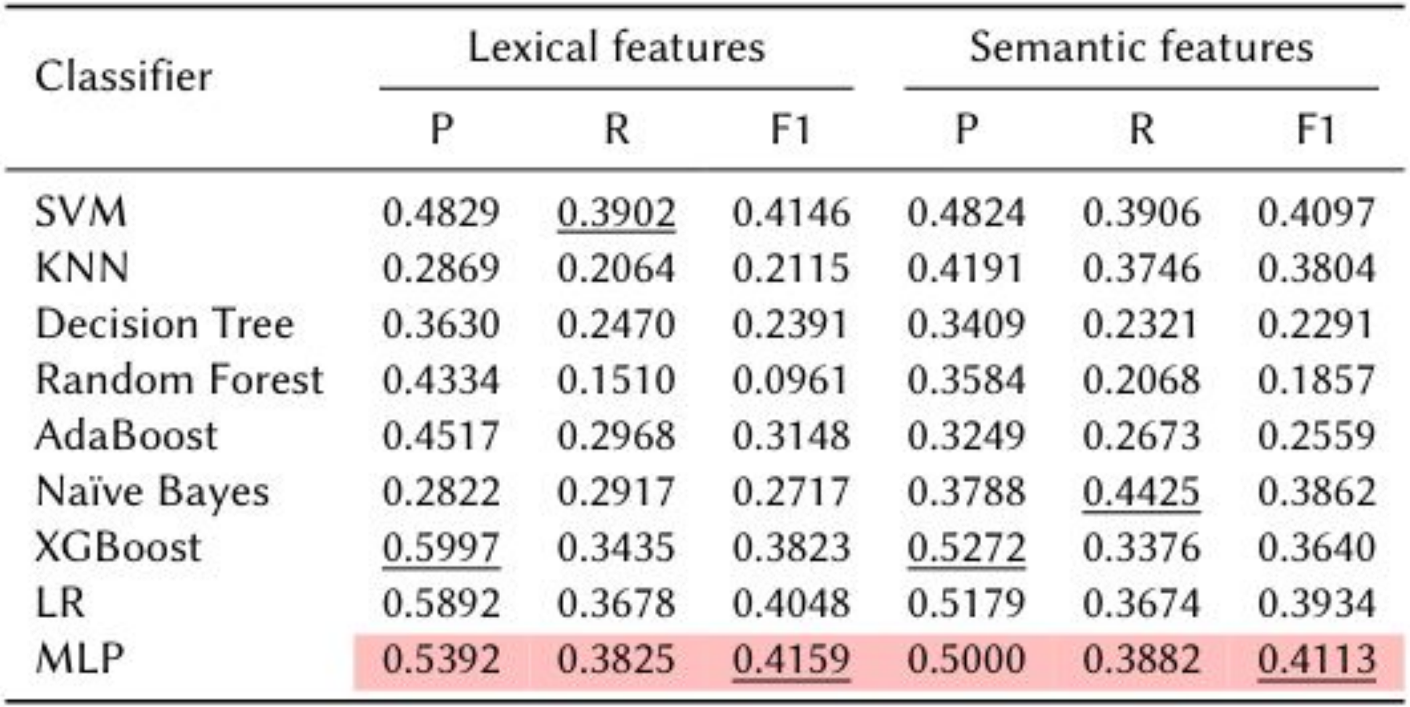
- Dataset published by Parikh et al (2021):
  - 60 judgments from Indian Supreme Court
  - 10,024 sentences
  - 7 rhetorical labels: facts, ruling by lower court, argument, statute, precedent, ratio of the decision, ruling by present court
- Baselines:
  - TF-IDF vectors (lexical feature set)
  - SBERT dense vectors (semantic feature set)
- 8 classic classifiers (knn, SVM,...) + 1 NN
- 26 models in total



Research Questions

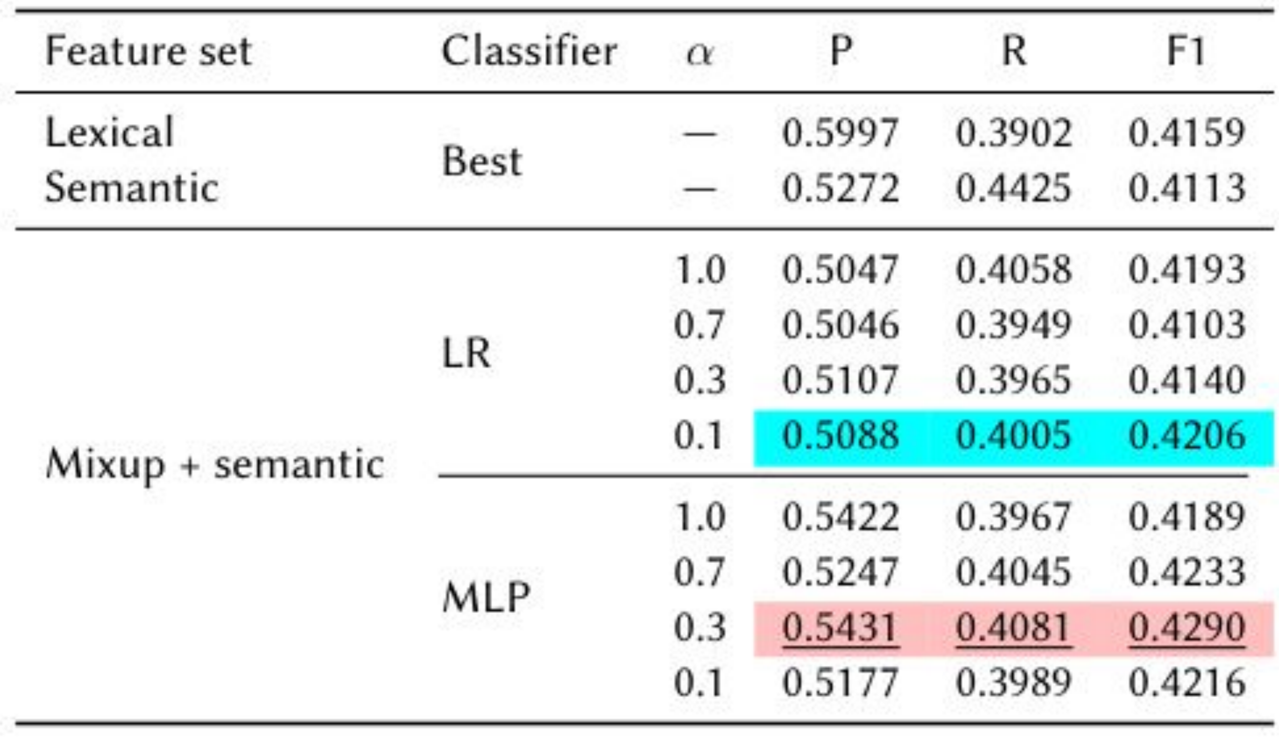
- Are SOTA models pre-trained with general-domain corpus able to produce useful representations for the task of rhetorical role labeling from Indian legal documents?
- Is the augmented data generated by the Mixup method able to improve the performance of semantic models?

Baseline results



Classifier	Lexical features			Semantic features		
	P	R	F1	P	R	F1
SVM	0.4829	0.3902	0.4146	0.4824	0.3906	0.4097
KNN	0.2869	0.2064	0.2115	0.4191	0.3746	0.3804
Decision Tree	0.3630	0.2470	0.2391	0.3409	0.2321	0.2291
Random Forest	0.4334	0.1510	0.0961	0.3584	0.2068	0.1857
AdaBoost	0.4517	0.2968	0.3148	0.3249	0.2673	0.2559
Naïve Bayes	0.2822	0.2917	0.2717	0.3788	0.4425	0.3862
XGBoost	0.5997	0.3435	0.3823	0.5272	0.3376	0.3640
LR	0.5892	0.3678	0.4048	0.5179	0.3674	0.3934
MLP	0.5392	0.3825	0.4159	0.5000	0.3882	0.4113

Mixup results



Feature set	Classifier	α	P	R	F1
Lexical	Best	—	0.5997	0.3902	0.4159
Semantic		—	0.5272	0.4425	0.4113
Mixup + semantic	LR	1.0	0.5047	0.4058	0.4193
		0.7	0.5046	0.3949	0.4103
		0.3	0.5107	0.3965	0.4140
		0.1	0.5088	0.4005	0.4206
	MLP	1.0	0.5422	0.3967	0.4189
		0.7	0.5247	0.4045	0.4233
		0.3	0.5431	0.4081	0.4290
		0.1	0.5177	0.3989	0.4216

Conclusions

- In the task of role labelling in the legal domain, our results show a small difference between lexical and semantic features with a slightly better performance of the lexical model regarding F1
- Our proposal, the use of Mixup in the legal domain, is capable to boost the semantic features and allow the models to overcome the baselines
- For future work, we plan to exploit pre-trained models on legal documents

References

- V. Parikh, U. Bhattacharya, P. Mehta, A. Bandyopadhyay, P. Bhattacharya, K. Ghosh, S. Ghosh, A. Pal, A. Bhattacharya, P. Majumder, **AILA 2021: Shared task on artificial intelligence for legal assistance**, in: D. Ganguly, S. Gangopadhyay, M. Mitra, P. Majumder (Eds.), FIRE 2021: Forum for Information Retrieval Evaluation, Virtual Event, India, December 13 - 17, 2021, ACM, 2021, pp. 12–15
- N. Reimers, I. Gurevych, **Sentence-bert: Sentence embeddings using siamese bert-networks**, in: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, 2019.
- H. Zhang, M. Cissé, Y. N. Dauphin, D. Lopez-Paz, **mixup: Beyond empirical risk minimization**, in: 6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings, OpenReview.net, 2018.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. **BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding**. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.