Rebalancing Label Distribution while Eliminating Inherent Waiting Time in Multi Label Active Learning applied to Transformers

Maxime Arens , *Lucile Callebert, Jose G. Moreno and Mohand Boughanem*

Contact : maxime.arens@gmail.com

LREC-COLING 2024

Context

Active learning (AL) can be applied to transformers

→ Reducing human annotation cost

→ Interesting for Multi-Label (several labels to annotate)

→ Transformers are well suited for uncertainty-based AL

→ AL works often focus on devising the better AL strategy in a given setting

Problem

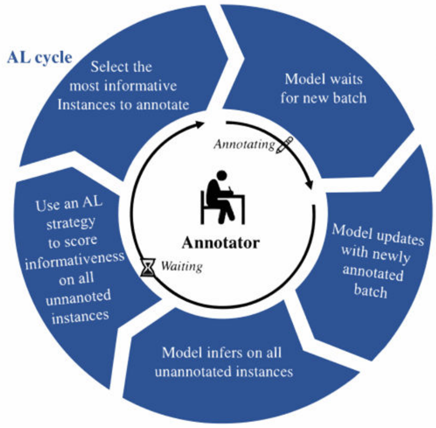
Applying AL to transformers in practical settings raises new challenges

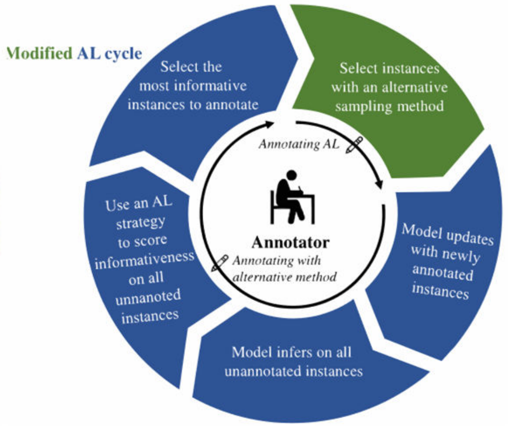
→ Some AL strategies sometimes perform worse than random sampling

→ It is difficult to predict which AL strategy will perform better on an unknown dataset

→ With transformers, the time between two AL cycles increases and leads to annotator idle time

Solution





Add an alternative sampling method to a classic AL workflow

→ random sampling

→ stale sampling : using score from previous AL cycles

→ Eq_labels (own) focuses on annotating rare labels

$$\text{Eq\_labels}(t)_i = \overbrace{c_i}^{\text{Confidence instance } i} . \sum_{j=1}^{q} \underbrace{y_i^j}_{\text{Prediction } i \text{ is labeled } j} softmax\Big(\overbrace{\frac{\max(L(t))+L(t)^j}{2.L(t)^j}}^{\text{Rarety of labels } j}\Big)$$

Findings

→ Adding alternatives sampling methods often leads to increased performances (percentage increase micro-F1 from classical AL)

→ Strategies that performed worse benefit the most from alternative sampling

	Method	Random		Stale		Eq_label	
	Model	distilBERT	distilRoBERTa	distilBERT	distilRoBERTa	distilBERT	distilRoBERTa
Strategy	ML	+22.807	+8.633	+13.073	-0.691	**+23.599**	**+11.051**
	MML	+7.188	+0.671	+2.200	-0.447	**+8.704**	**+3.891**
	CMN	-1.318	-1.570	-0.132	-3.947	**+3.998**	**+0.806**
	MMU	+2.111	-4.289	+0.528	-2.649	**+3.254**	**+0.210**
	LCI	+14.764	**+7.427**	+1.990	-1.294	**+17.958**	+5.319
	CVIRS	+6.342	**+2.585**	**+9.276**	-2.139	+3.975	+1.381

Eq_labels reduces label imbalance

→ Over classical AL and other alternatives methods

→ Over full dataset distribution on three out of four datasets (Mean Imbalance Ratio)

Dataset	Method / Model	Classic-AL	Random	Stale	Eq_label	Full
Jigsaw_toxic	distilBERT	6.681	6.846	5.776	**4.886**	9.537
	distilRoBERTa	6.782	6.869	6.548	**5.874**	
Go_emotions	distilBERT	12.945	13.035	15.065	**10.017**	12.661
	distilRoBERTa	14.238	12.719	15.111	**10.690**	
EUR_Lex	distilBERT	54.560	37.422	53.409	34.826	**25.918**
	distilRoBERTa	37.745	34.667	38.288	30.048	
UNFAIR-ToS	distilBERT	5.945	5.018	5.781	**2.795**	3.301
	distilRoBERTa	5.960	5.972	7.286	**2.894**	

Conclusion

→ Eliminate annotator waiting time

→ Reduce differences in performances between different AL strategies

→ Rebalance label distribution during annotation

→ Provide a sampling method that improves performances over six AL strategies for two models and four datasets

Synapse

IRIT

in