ASC: Aggregating Sentence-level Classifications for Multi-label Long Text Classification

Takahiro Komamizu (Nagoya Univ., Japan) taka-coma@acm.org

Multi-label Long Text Classification (MLLTC)

A special case of text classification that
- **Longer text** (than the length limit of classification models; esp. pre-trained language models (PLM))
- **Multiple labels on the text** (imagine the fine-grained labels like topics of your papers)

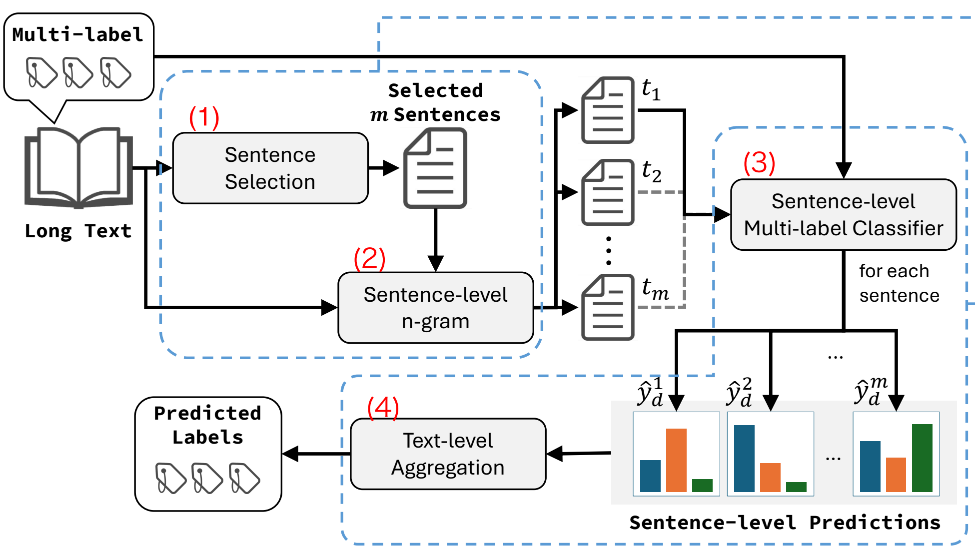
Challenges:
- Handling long text within PLM input length limits
- Predicting multiple labels, especially for tail classes (caused by long-tail distribution)


ASC: Aggregating Sentence-level Classifications for Multi-label Long Text Classification

Topic Labels
- Long Text Classification
- Multi-label Classification
- Prediction Aggregation
- Sentence-level Classification
- Pre-trained Language Models
- Extractive Summarization
- Sentence-level n-grams
- Class Imbalance
- Efficient Training

etc.

ASC: the Simple-yet-Effective Proposed Method



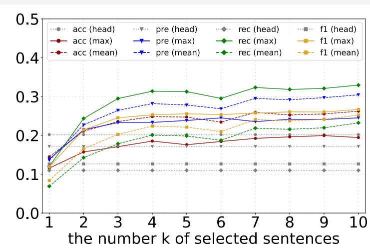
Segment document carefully.
- Extractive summarization (e.g., TextRank[17]) to select k key sentences
- Sentence-level n-gram for context reconstruction

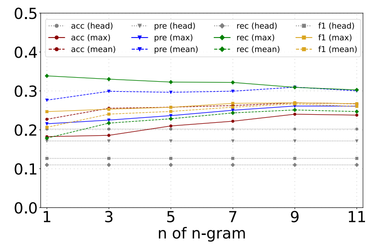
Estimate local labels and aggregate to global labels.
- PLM-based classifier for sentence-level classification
- Aggregation for each label by Max or Mean (or any)

Experimental Evaluation on Reuter-21578 and EURLex-57K with Macro-averaged Metrics

Method	Reuter-21578						EURLex-57K					
	k	n	Acc ↑	Pre ↑	Rec ↑	F1 ↑	k	n	Acc ↑	Pre ↑	Rec ↑	F1 ↑
In-Length Limit Methods (Sentences to fit in length limit)												
DistilBERT-head	-	-	.843	.382	.296	.322	-	-	.202	.171	.110	.127
DistilBERT-ES	5	-	.838	.382	.300	.325	5	-	.153	.153	.097	.111
DistilBERT-ES	10	-	.834	.383	.293	.319	10	-	.162	.158	.099	.114
DistilBERT-ES	20	-	.836	.389	.289	.317	20	-	.169	.162	.101	.117
Long Text Handling Methods (Use as much sent. as possible)												
DistilBERT-Rand [20]	-	-	.858	.497	.386	.419	-	-	.224	.221	.163	.179
DistilBERT-TR [20]	-	-	**.860**	.488	.366	.397	-	-	.225	.230	.167	.184
ToBERT [19]	-	-	.850	.478	.377	.406	-	-	.166	.137	.071	.087
LongFormer [2]	-	-	.850	.450	.359	.384	-	-	.099	.131	.086	.099
Proposed Methods (Selective plus contexts)												
ASC-mean	9	11	.852	**.548**	.457	.486	8	9	**.274**	**.310**	.247	.263
ASC-mean w/o ES	-	11	.844	.546	.453	.480	-	9	.269	**.310**	.251	.267
ASC-max	9	11	.824	.539	**.537**	**.524**	8	9	.236	.262	**.318**	**.273**
ASC-max w/o ES	-	11	.819	.522	.530	.516	-	9	.240	.261	.309	.270

Analyses on EURLex-57K





- The careful selection of sentences is effective.
- The contextualization of sentences is effective.
- Single representation is not effective in MLLTC.

- # training samples increases by sentence selection, and this leads higher training cost.
- ASC still suffers from the class imbalance issue.