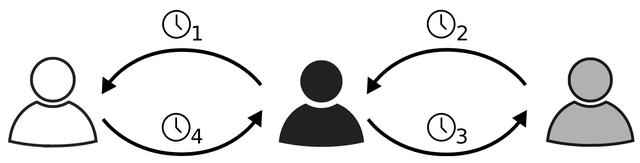
Putting Context in Context: the Impact of Discussion Structure on Text Classification

Nicolò Penzo, Antonio Longa, Bruno Lepri, Sara Tonelli, Marco Guerini

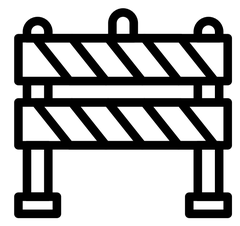
Discussion Chain

Local Discussion Network



Goal

Classify content of the last message

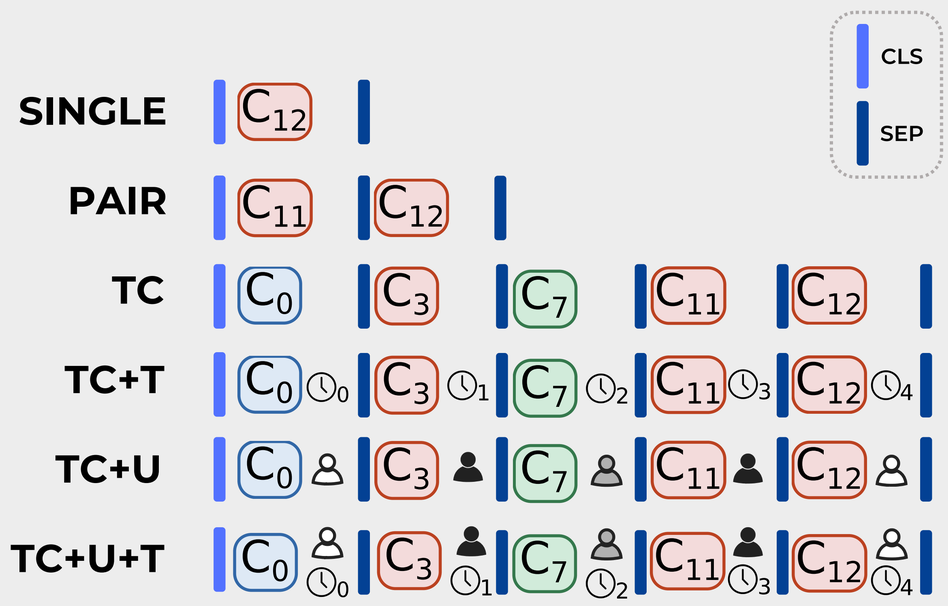


Constraint

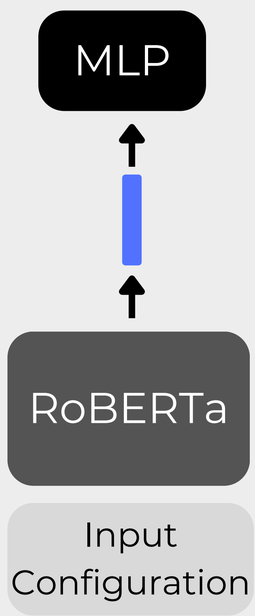
Privacy-preserving

Profiling-preserving

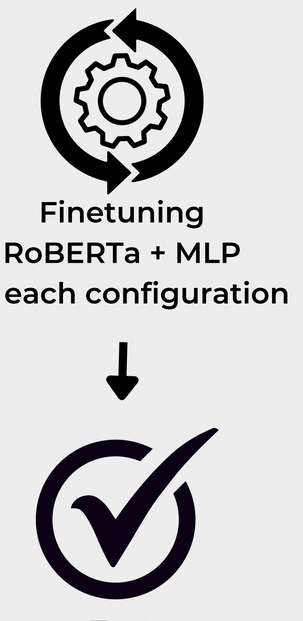
Input Configurations



Model

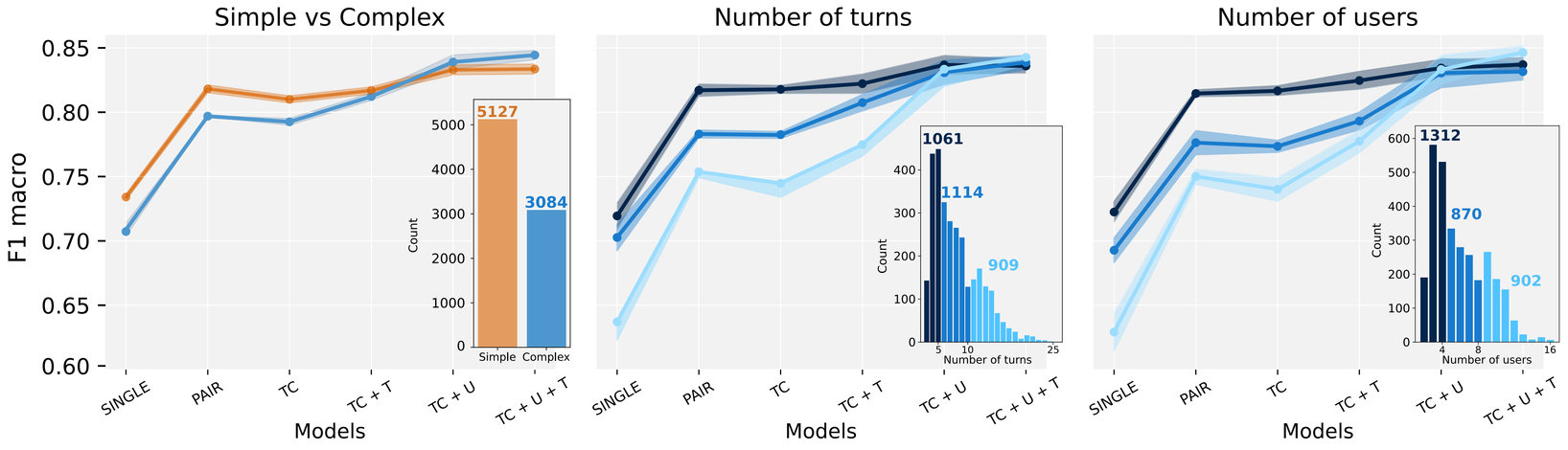


Pipeline

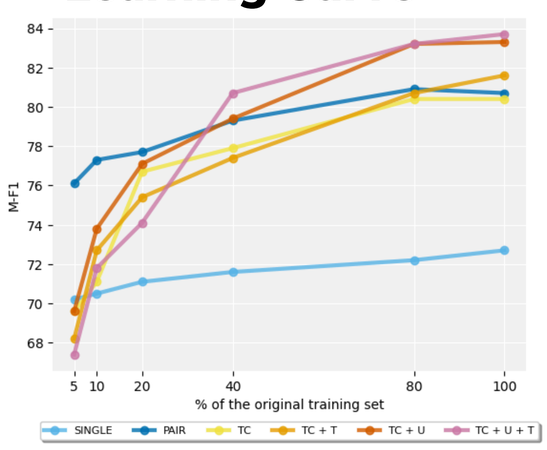


- Three text classification datasets
- Analysis of the correlation with number of turns
- Analysis of the correlation with number of users
- Learning curve analysis on the biggest (100k items in training)

Correlation with Discussion Complexity



Learning Curve



FINDINGS

- **Full linguistic context alone worsens** or does not significantly improve the results with respect to the non-contextual baseline.
- With **extra-linguistic context**, the performance **improves**, especially with the contribution of structural context.
- Analysis on the learning curve shows that **results strongly depend on the amount of training data**.
- **Extra-linguistic context** makes results more **robust** across discussion networks of different lengths and more or less active users.
- **Transformer-based models** are able to **embed structural features**, given in input to the model in the form of **simple natural language statements**.

EACL 2024

Reference:
Nicolò Penzo, Antonio Longa, Bruno Lepri, Sara Tonelli, and Marco Guerini. 2024. Putting Context in Context: the Impact of Discussion Structure on Text Classification. In Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics (Volume 1: Long Papers), pages 1793–1811, St. Julian's, Malta. Association for Computational Linguistics.

PAPER

POSTER

REPO

Nicolò Penzo
npenzo@fbk.eu
@penzo_nicolo
nicolopenzo.github.io

Affiliations
University of Trento, Trento (Italy)
Fondazione Bruno Kessler, Trento (Italy)