Can Unsupervised Knowledge Transfer from Social Discussions Help Argument Mining?

SUBHABRATA DUTTA[1], JEEVESH JUNEJA[1], DIPANKAR DAS, TANMOY CHAKRABORTY

ACL 2022
22ND – 27TH MAY | 60TH MEETING | DUBLIN

KEYWORDS

ACI. Argument Component Identification, refers to the task of identifying argument components(claims, premises etc.) at the token level.

RTP. Relation Type Prediction, refers to the task of predicting the type of relation between two related components, e.g. attack, support, etc.

Discourse Markers. These are words or phrases whose main role is to coherently tie together the statements in a discourse. For e.g., "I think,", "however", "if", "so", "imo", "tldr" etc.

Threads. We consider social discussions from the *ChangeMyView* reddit forum. The data is in the form of **threads**, where each thread is a sequence of **posts** by various users.

We use the **LongFormer** model with a sequence length of 4096 tokens to utilise long contexts of social discussions efficiently.

A SAMPLE THREAD

We investigate both **post/comment**-level and **thread**-level contexts for our tasks. Claims and Premises annotated in a CMV thread are shown below. We use the annotated threads from *CMV-Modes* data released in **AMPERSAND**.

EMTs, SAR, firefighters, police, etc. should receive "military discounts". For those of you who don't know, it's common (at least in the US) for businesses, transit agencies, etc. to give small discounts to military veterans to thank them for their service. It seems that medical responders (even hospital staff, actually) and other emergency services do more good for society than soldiers and that such discounts should be given to them.

> *other emergency services do more good for society than soldiers and that such discounts should be given to them.*

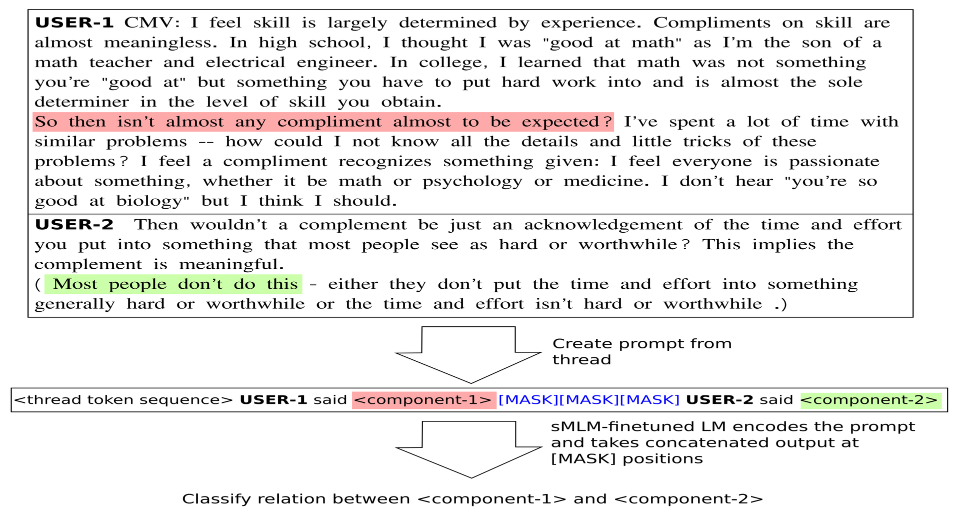
What about teachers? Scientists? Doctors? Farmers? There are plenty of professions that do great things for society but people aren't going around thanking them with discounts.

In that case, what makes soldiers more worthy than others?

I don't believe I ever argued that they were. Your premise is that these professions "do more good" than soldiers, and I'm asking why you don't want to extend the benefits to other critical pillars of our society.

PROMPT FINE-TUNING

For our RTP task, at the **thread**-level, we try prompt-based fine-tuning, mean pooling fine-tuning(**mp**)(pooling over the components between which relation is to be predicted). Below we show how we generate our prompts.



SELECTIVE MASKED LANGUAGE MODELING

We define a novel selective masked language model(**sMLM**) task. This task involves MLM training, but we only mask a selected set of discourse markers. It is carried out on a large corpus of un-annotated data from CMV forum. We use the *Winning Args* dataset collected by Tan et. al. 2016 as provided in ConvoKit.

u/DurianMD:
CMV: Religion is not violent or not violent, its followers are. So, my belief is that while religion can inform the views of people, it is far more likely that religion will be used to justify actions that would have been executed any way. I think that most Jewish people don't want to stone adulterers and most Muslims don't want to stone non believers.

u/recycled_kevlar:
Your stance relies on the assumption that religion has no influence on the actions of its followers beyond the superficial. Yet something must exist that allows this pattern to occur. Ill narrow it down to religion or culture. So, you are correct if you assume the culture dominates the religion, and you are incorrect if the reverse is true. With this in mind, I think its safe to assume the truth is somewhere in between, with both the religion and the culture somehow influencing the unrest we see.

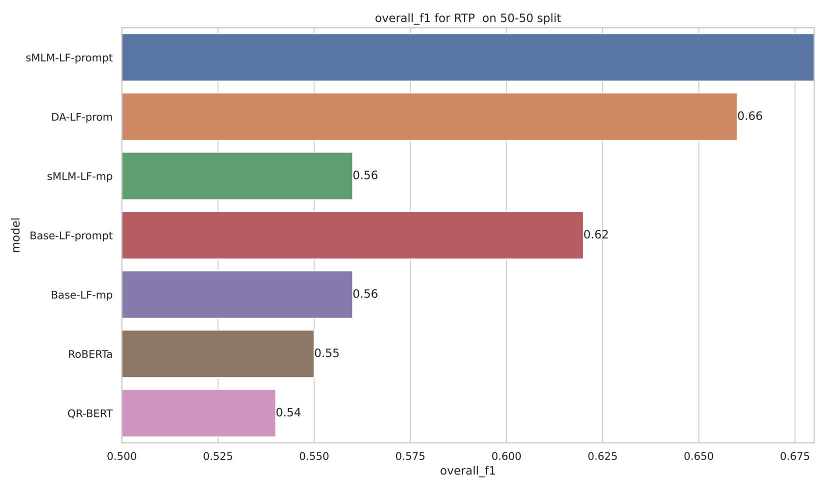
u/DurianMD:
I suppose I was taking a harsh stance when I assumed that religion had no effect on behavior, when it obviously does. I still think the culture dominates religion to a great extent, however I cannot ignore that religion does have an effect on culture to some extent.

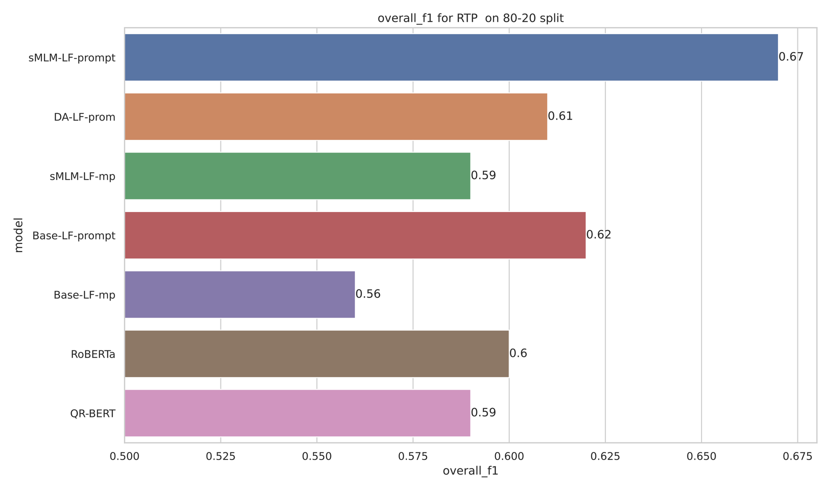
CENTRAL CONJECTURE

Our main conjecture is that **sMLM**-fine-tuning results in significant performance gains for argument mining tasks, over MLM and domain-adaptation.

It helps us utilise un-annotated data from social discussions, in the lack of annotated data. It biases the model to utilise long thread level contexts efficiently during fine-tuning for final task.

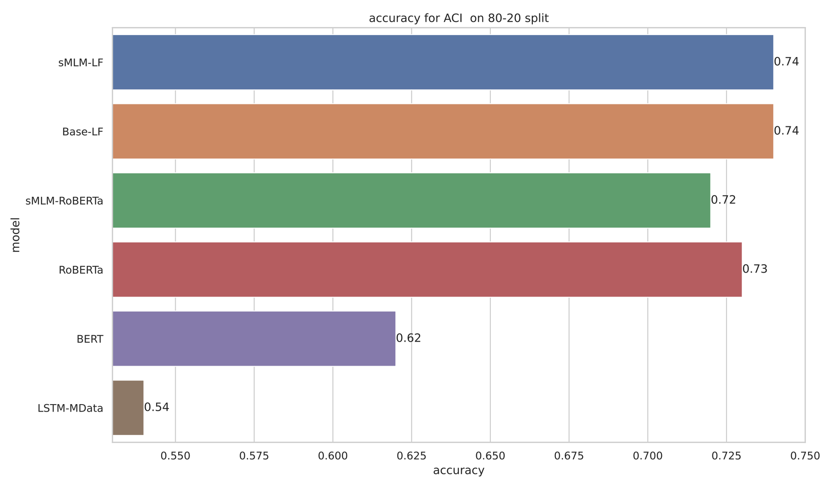
EVIDENCE-I: RTP

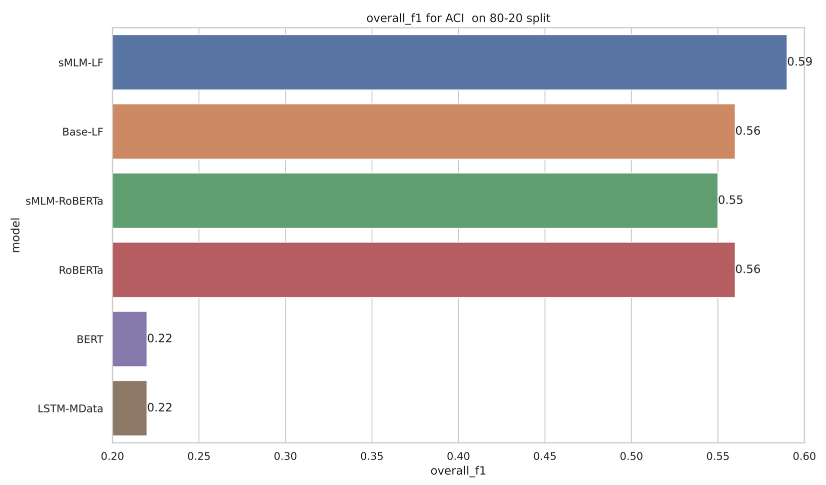




1. Our **sMLM-LF** variant, with prompt based fine-tuning performs better than all models on both splits. Even better than the Domain Adapted(**DA**) LongFormer.
2. **sMLM** leads to much consistent and larger benefits when used with prompt based finetuning rather than mean pooling strategy.
3. **RoBERTa** and **QR-BERT** models, which are trained on **component**-level, perform worse than models fine-tuned with the context of entire thread.

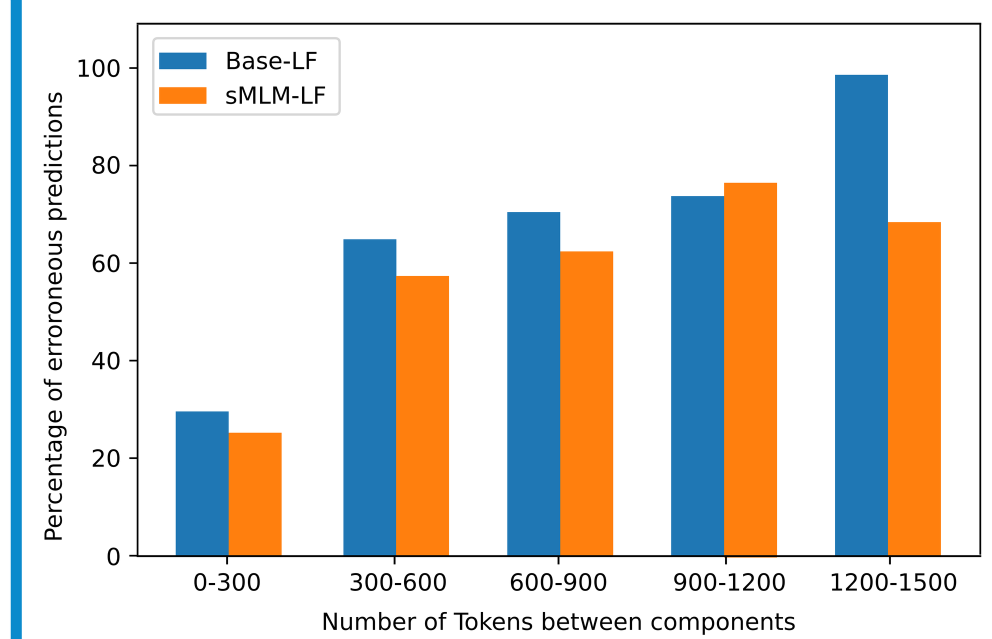
EVIDENCE-II: ACI





1. We see a significant increase in accuracy as we move from LSTM-based models to BERT, probably because of large number of parameters and pretraining data for transformers.
2. RoBERTa performs much better than BERT in terms of F1 scores, and **sMLM** trained Longformer, performs better than all of these.
3. Observing Class-wise(Claims, Premises etc.) scores, reveal that **sMLM** training leads to larger benefit for classes which need more context for prediction.

ERROR ANALYSIS



1. Error Rates vary proportionally with distance, but **sMLM-LF** consistently yields lower error rates than **Base-LF**.
2. For components which are far apart, we see significantly lower error rates for **sMLM**(last bars).