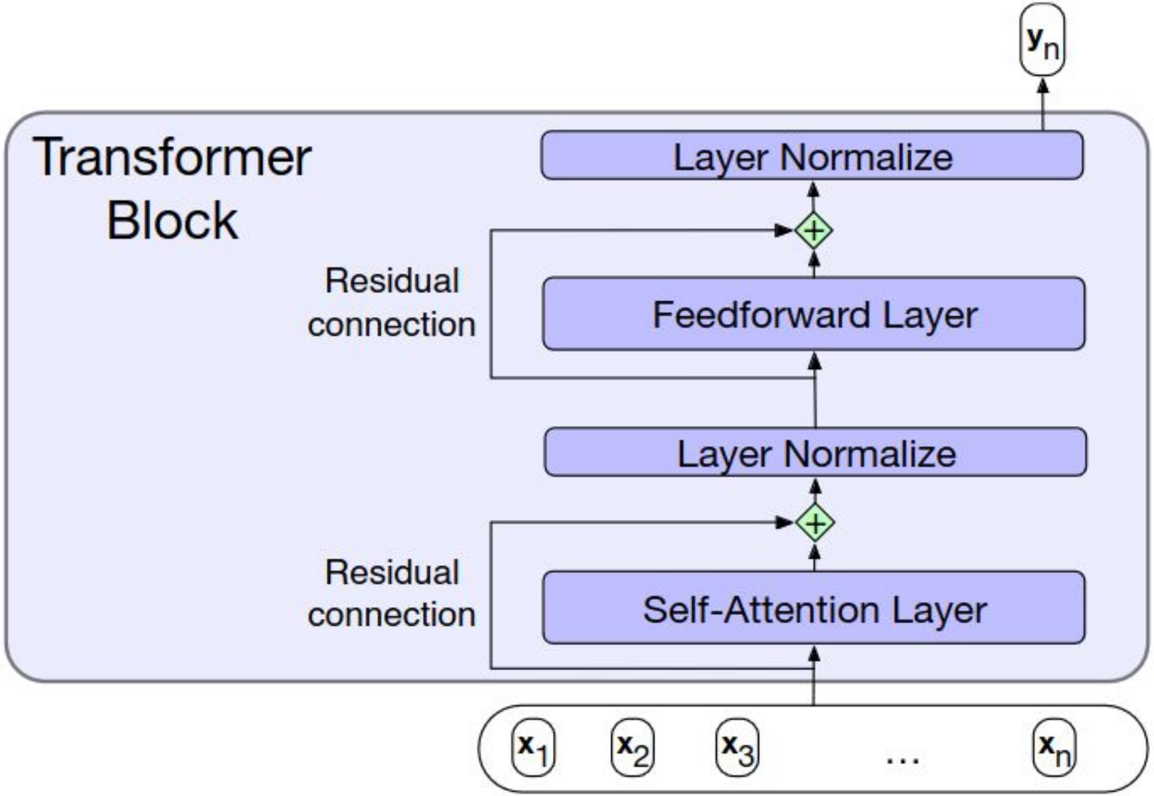
From word embeddings to Language Models

Segun Aroyehun

Language models

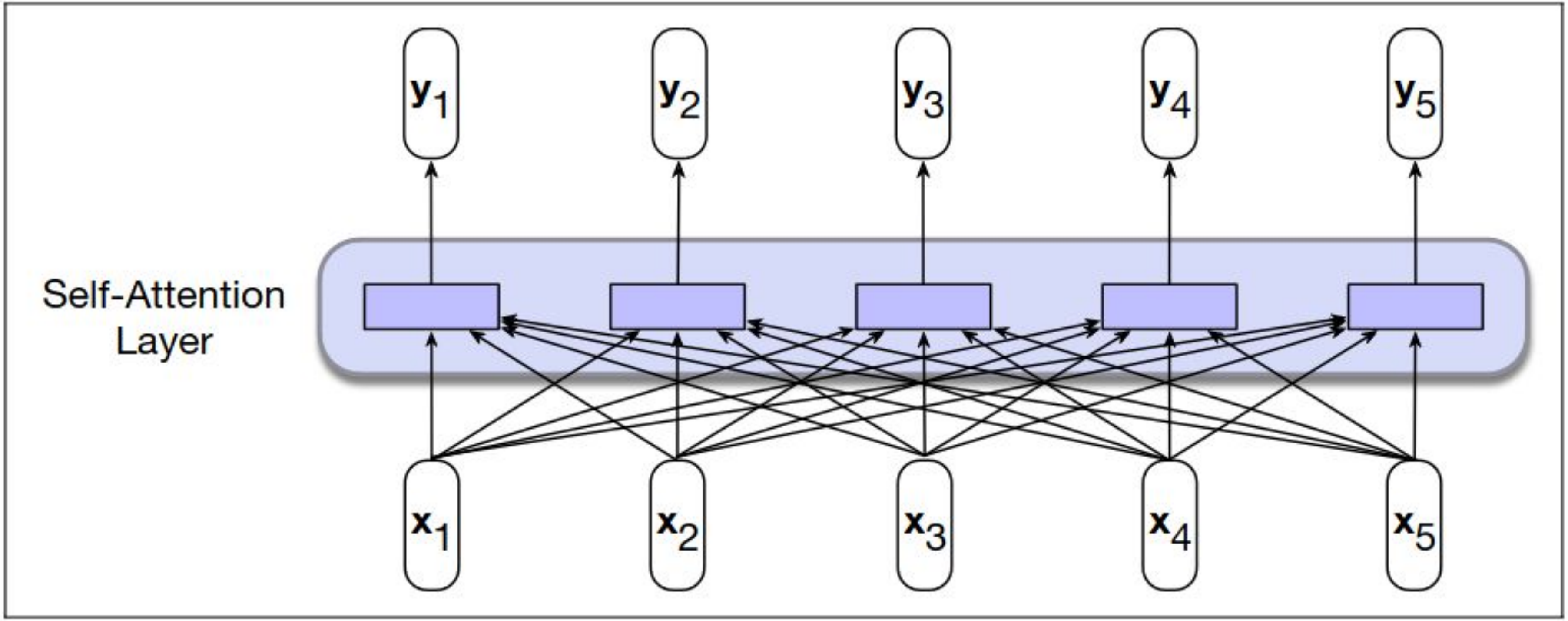
- A model that learns a statistical pattern of occurrence of sequence
- Language models as next word/sentence prediction
  - E.g. autocompletion on mobile keyboard
- Recent language models are based on the Transformers architecture

Transformers



Speech and Language Processing. Daniel Jurafsky & James H. Martin (2023)

Self-Attention



Speech and Language Processing. Daniel Jurafsky & James H. Martin (2023)

BERT: Brief Introduction

Bidirectional Encoder Representations from Transformers (BERT)

- Transformers are neural networks which use attention mechanism
- A “large” language model (LM)
  - Base with 12 Transformer layers
  - Large with 24 Transformer layers
- Trained on large collection of text: Wikipedia and BookCorpus
- With several specialized hardware GPUs/TPUs
- Examples of LMs based on Transformers: RoBERTa, XLNET, ELECTRA, GPT-4

BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding (Devlin et al., 2018)

Word Vectors vs. Contextual LM

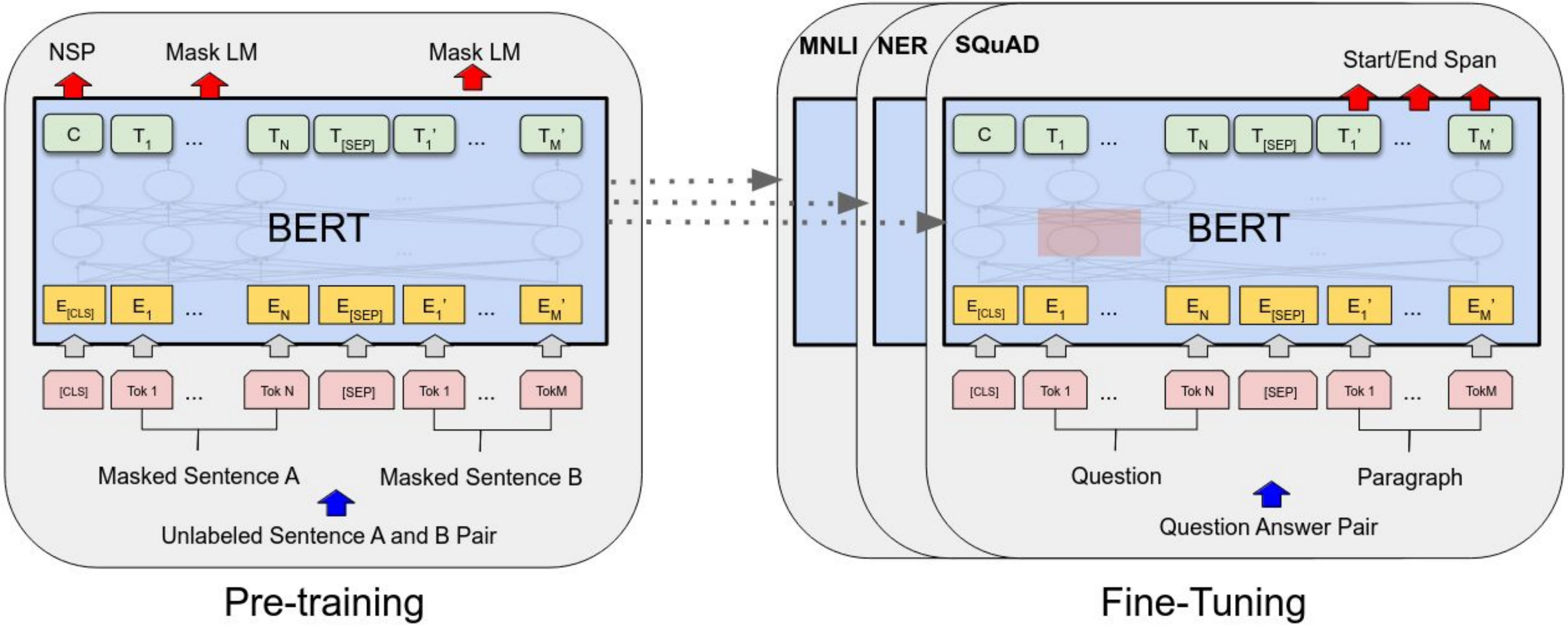
- Word vector: A representation per word e.g. word2vec, GloVe, fasttext
- Contextual LM: A representation per word <u>in context</u>

Richer semantic representation

One model, many applications

- Model sharing and reuse
- 🤗 HuggingFace library and hub
- From a representation to several applications:
    - Topic classification
    - Natural language inference
    - Question answering
    - Sentiment analysis
    - Emotion detection
    - Machine translation

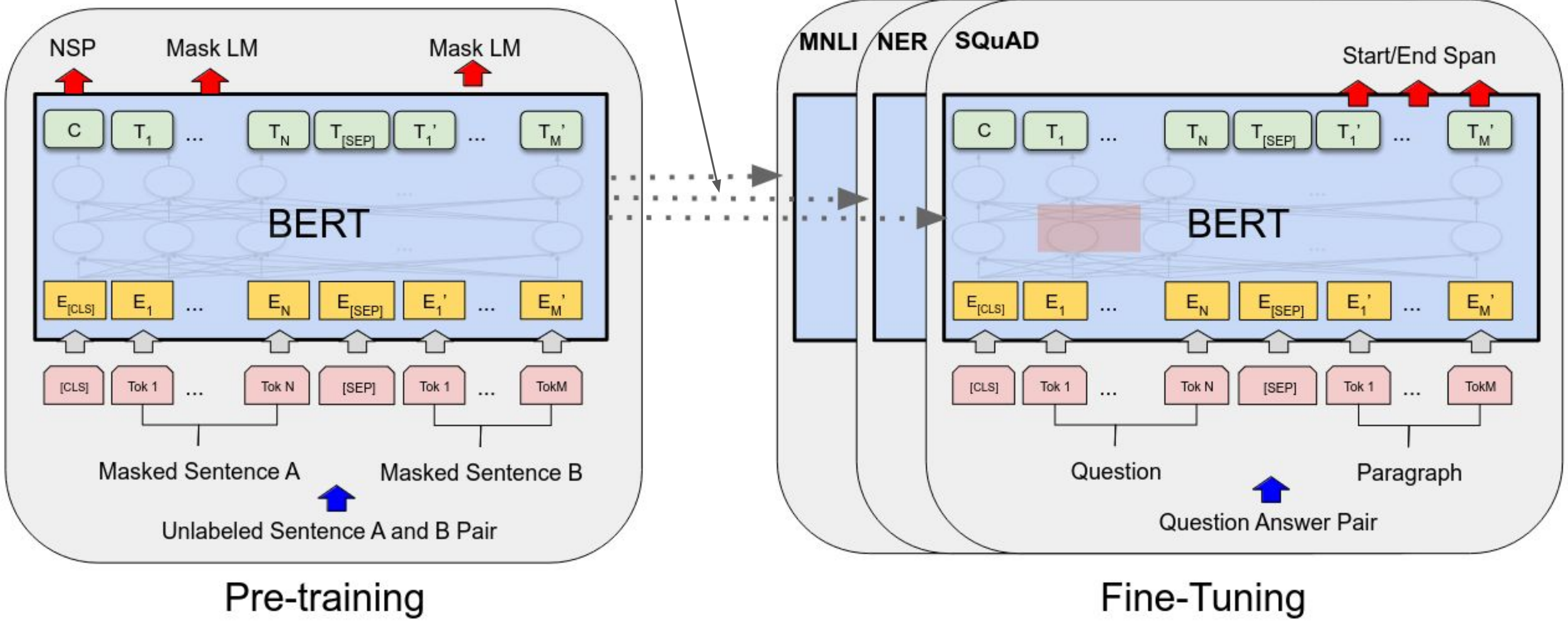
Pre-training and Fine-tuning



BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding (Devlin et al., 2018)

Problem: Domain gap

Solution: Pre-train on task-specific data

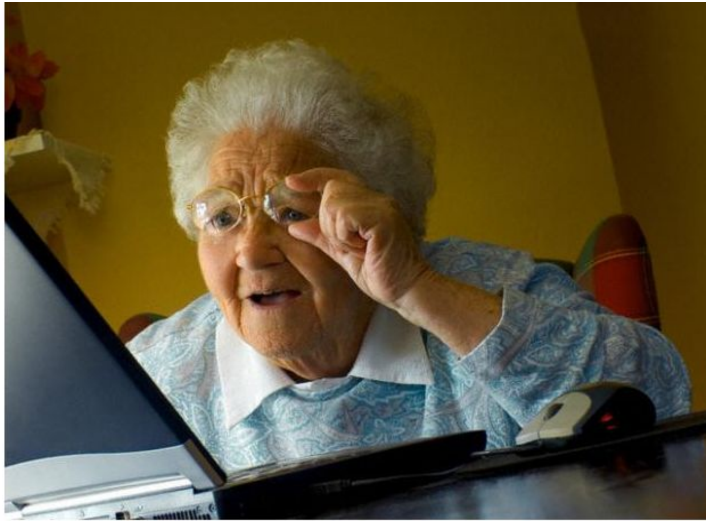


Examples: SciBERT, TwitterRoBERTa, BERTweet

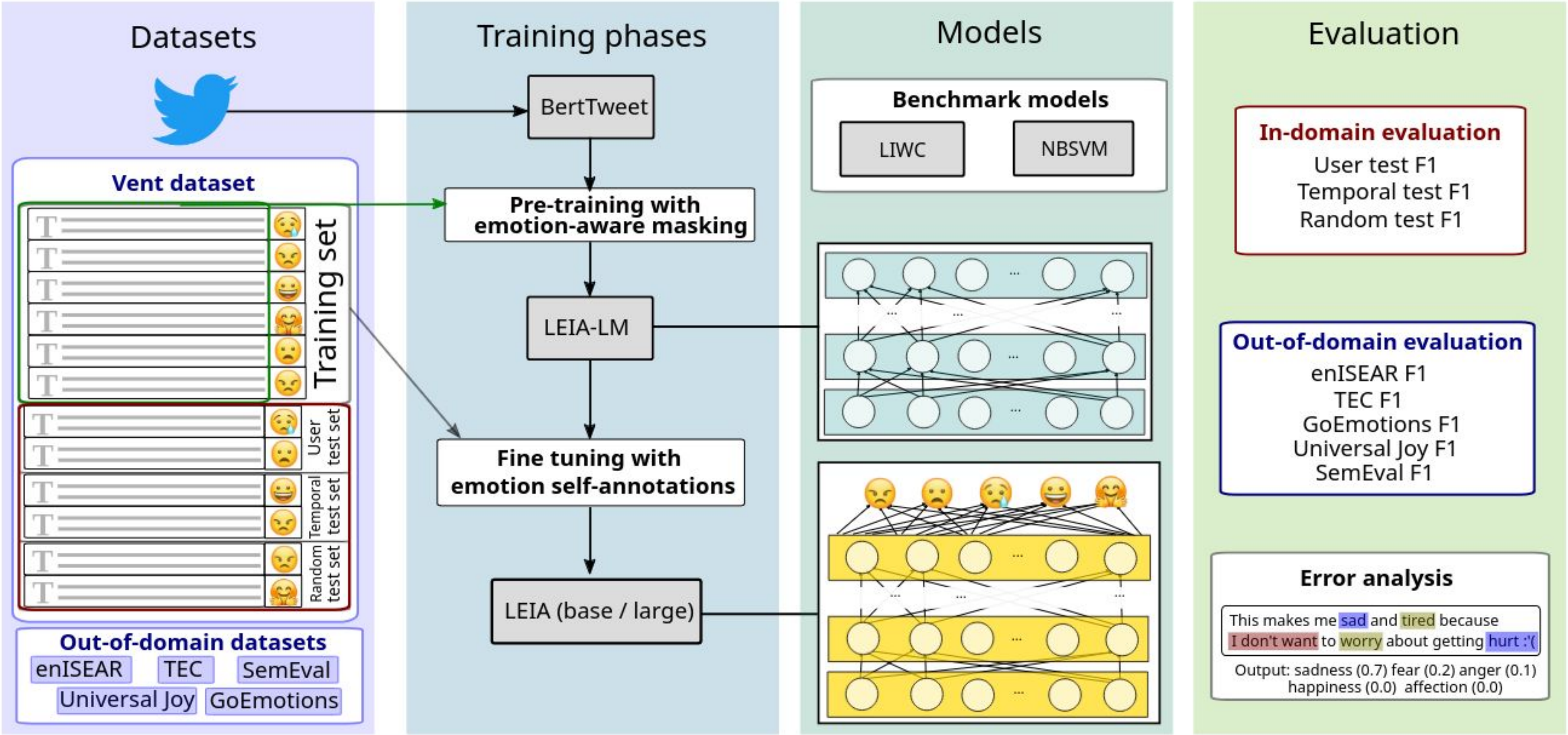
LEIA: Linguistic Embeddings for the Identification of Affect (Aroyehun et al., 2023, preprint)

Emotion identification in text

Existing models are trained mainly on datasets annotated by readers than writers

LEIA: An overview



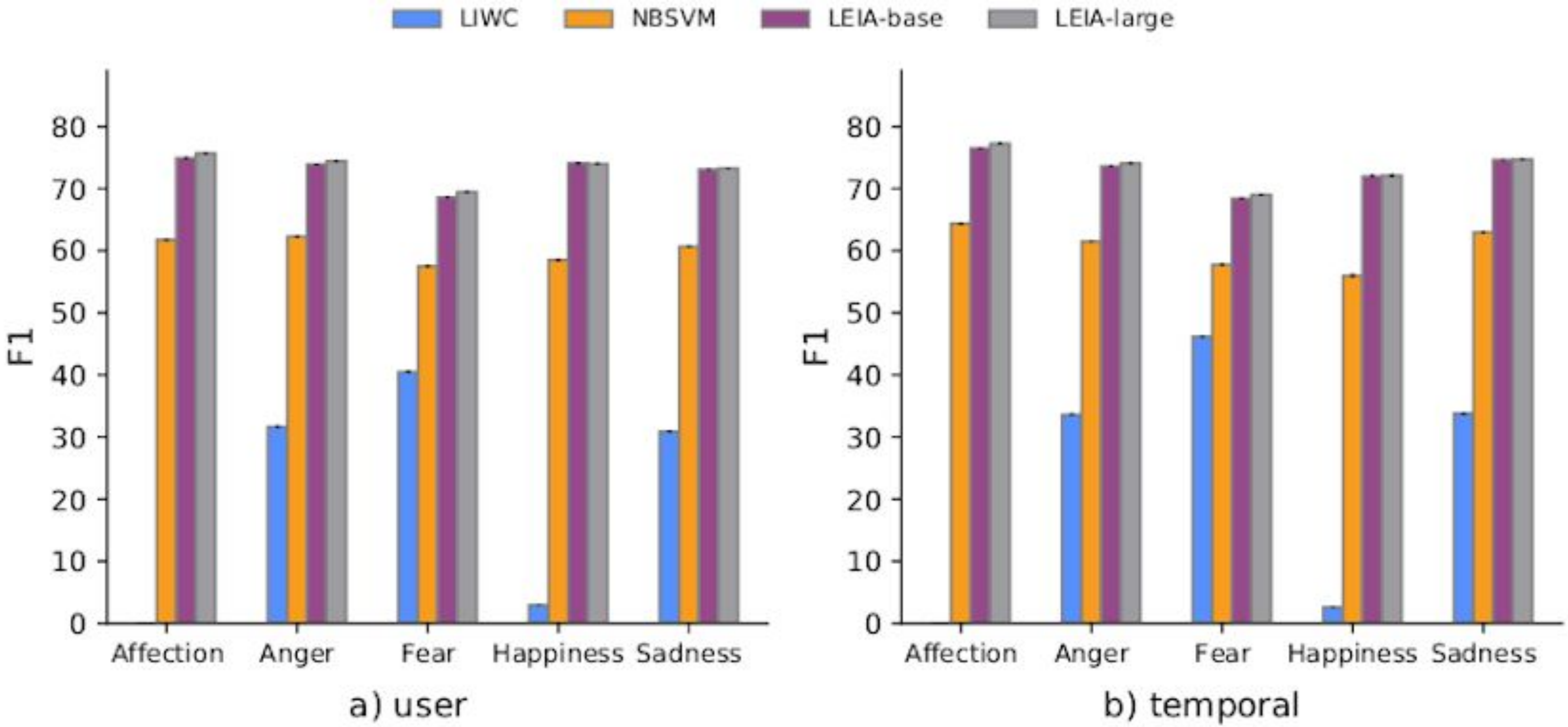
Datasets

- Vent as a source of self-annotated dataset
- Sharing emotions at scale: The Vent dataset (Lykousas et al., 2019) : A dump of 33M posts
- Labels : Affection, Anger, Fear, Happiness, Sadness
- In-domain evaluation on user, temporal, and random splits
- Out-of-domain evaluation: Universal joy, GoEmotions, enIsear, TEC, and Semeval
- Out-of-domain (OOD) label groupings exclude Affection

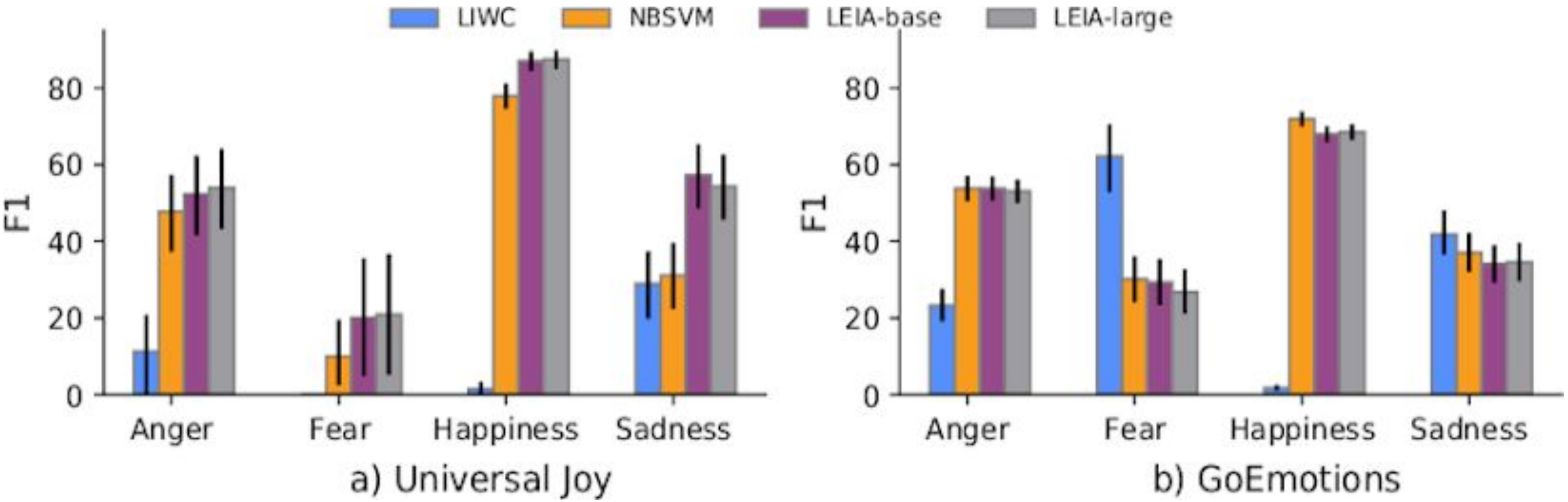
Training the LMs

- We experiment with two LMs : BERTweet-base and BERTweet-large
- Adaptation by pre-training with selective masking of emotion words on unlabeled data
- Fine-tuning on labeled data
- We evaluate on unseen in-domain and out-of-domain data in comparison with two other models

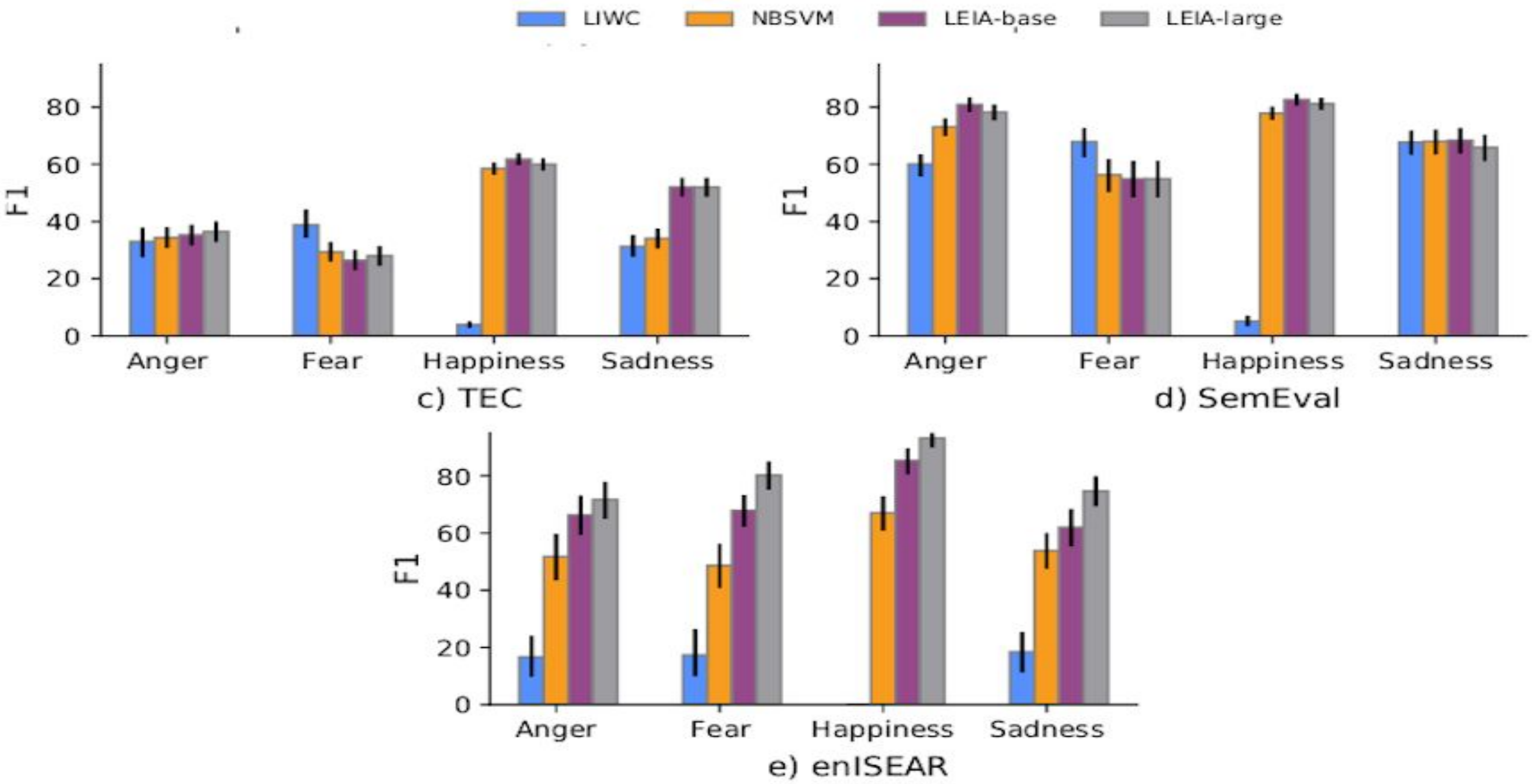
Results in-domain



Results out-of-domain I



Results out-of-domain II



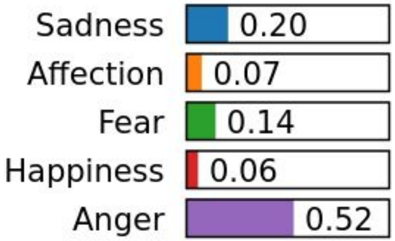
LEIA is on 🤗 HuggingFace hub

Two models in two sizes:

- Adapted LM (base and large)
- Emotion classification model (LEIA-base and LEIA-large)

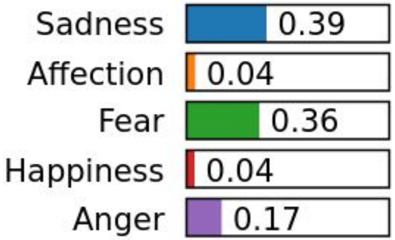
https://huggingface.co/LEIA

Sample model prediction and attributions



I felt |mask| because one of my children lied. I taught them not to, but natural sin came out and I was surprised to find myself so cross.

Actual label: Anger



I felt |mask| when my house from a few years ago got damaged by burst pipes and flooded internally. Most of our belongings were destroyed and we had to move out for nearly 18 months while the house dried out and was refurbished. It was an awful lot of hassle and upheaval.

Actual label: Sadness

Summary

- Semantic Differential
  - Connotative vs. denotative meaning
  - Three dimensions of meaning: Evaluation, Potency, and Activation
- Word embeddings
  - Cosine distance between vectors as a measure of similarity
  - Latent semantic analysis as a solution to the curse of dimensionality
  - Distributed representation
  - Distributed dictionary representation
- Language models
  - Pre-training and fine-tuning
  - Adaptation to specific tasks/domains before fine-tuning
  - LEIA as an example of adaptation for emotion classification

Thank you for listening!