Modeling user context for valence prediction from narratives

Aniruddha Tammewar[1], Alessandra Cervone[1], Eva-Maria Messner[2], Giuseppe Riccardi[1]
[1]Signals and Interactive Systems Lab, University of Trento, Italy
[2]Clinical Psychology and Psychotherapy, University of Ulm, Germany

UNIVERSITY OF TRENTO - Italy
Department of Information Engineering and Computer Science

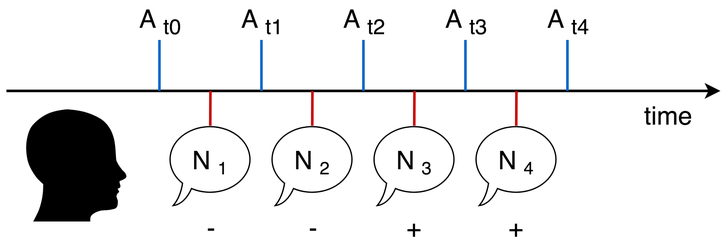
SiS LAB Signals & Interactive Systems

Introduction

- **Valence** indicates the **emotional value** (on a scale from positive to negative) associated with an event, situation or object.
- Automated prediction of valence may provide crucial information for mental healthcare
- Role of context in the task of valence prediction?

The ULM state of mind dataset (USoMS)

- Personal narratives with self-reported affect information
- Participants recount two negative and two positive events from their life for 5 mins
- Self assessment of valence using 10 point Likert scale, at five stages
- Scores grouped into three classes: Low (0-4), Medium (5-7) and High (8-10)



Data collection process: N1, N2 are Negative narratives; N3, N4 are positive narratives; while At0-4 are self reported affect scores

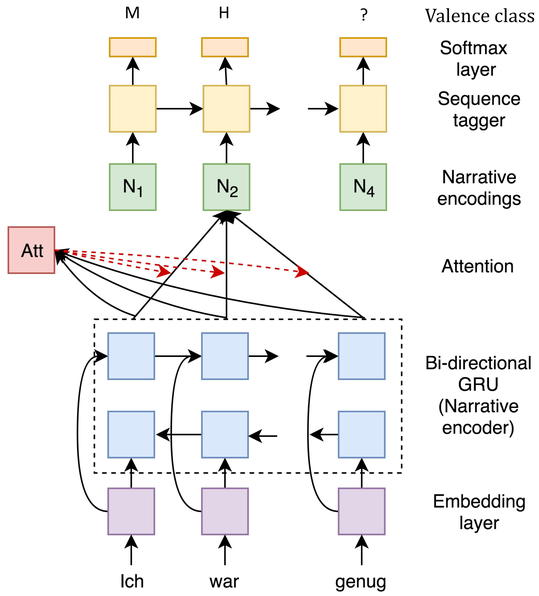
- A part of USoMS was used in the INTERSPEECH 2018 Computational Paralinguistics Challenge (ComParE)
  - Task: automatically predicting self-reported affect
  - Input: 8 seconds speech fragment
  - Acoustic features

Motivation

- Context provided by the full history of the:
  - Current narrative
  - All previous narratives by the same individual
- ComParE challenge does not allow exploring context
- Generalizability: utilizing solely the textual information

Methodology

- **Incorporate context :**
  - Feature Engineering
  - ML and DNN architectures



DNN based sequence tagging architecture

Are previous narratives recounted by an individual useful for predicting his/her current mental state?

Experiments and Results

Model	Narratives used	Features	Accuracy
Linear SVM	N_t	μ word emb	55.5 ±5.0
	N_t	μ word emb, pol	57.8 ±4.8
	N_t	tf-idf, pol	57.8 ±4.5
	N_{t-1} , N_t	tf-idf, pol	59.7 ±5.9
biRNN + attn	N_t	word emb, pol	58.2 ±6.8
Encoder (biRNN + attn) + RNN (context pair)	N_{t-1} , N_t	word emb, pol	**62.4** ±8.7
Encoder (biRNN + attn) + RNN (sequence tagging)	N_0 ,..., N_t	word emb, pol	61.8 ±6.4

- Previous context seems to be helpful for valence prediction
- In general Neural models outperform Linear SVM

Discussion

Distribution of attention weights on four (fragments of) consecutive narratives

- Sentiment carrying words and phrases (e.g. happy "*zufrieden*")
- **Emotion triggering concepts**
  - **people** (e.g. grandfather "*opa*")
  - **events** (e.g. high school exam "*abi*", trip after the high school exam "*abireise*")
  - **places** (e.g. university *"uni"*)
- **Disfluencies** ("*uhm", "ahm"*)

COADAPT
www.coadapt-project.eu