

LEADEMPATHY: An Expert Annotated German Dataset of Empathy in Written Leadership Communication

Didem Sedefoglu, Allison Lahnala,
Jasmin Wagner, Lucie Flek, Sandra Ohly

Leadership communication in the post-pandemic era

- Communication between leaders and followers has been shifting to virtual contexts
- Challenge to maintain trust and relationships (Marlow et al., 2017)
- Decreased richness of communication (Daft et al., 1987)
- Empathy as a key **leadership quality** and important aspect of **individual and organizational effectiveness** (Kock et al., 2019)
- AI technologies are increasingly being used in organizational practices (Holmström & Hällgren, 2021)



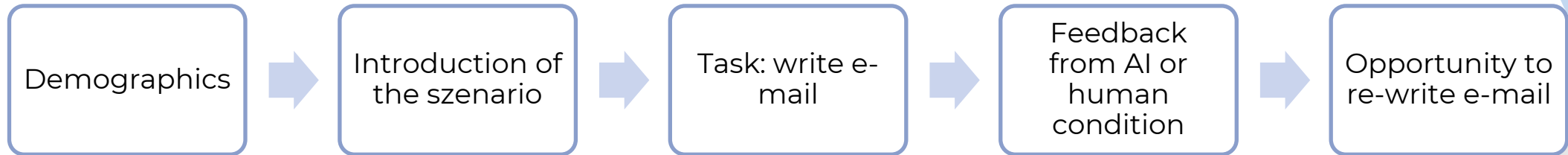
Related work in NLP

- Enhancing higher levels of empathy thorough AI-tools:
 - Peer-to-peer mental health support (Sharma et al., 2020)
 - Student peer reviews (Wambsganss et al., 2021)
- Empathy Detection Models (Buechel et al., 2018; Pérez-Rosas et al., 2017; Sharma et al., 2020; Wu et al., 2021; Wambsganss et al., 2022)
- Challenging to define and measure empathy
 - Reliance on broad conceptualizations of empathy (Rashkin et al., 2019)
 - Recently more theory-grounded approaches (e.g. Sharma et al., 2020, Wambsganss et al., 2021)

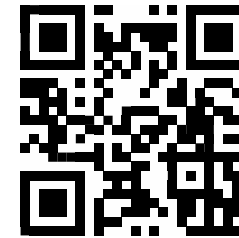
Our contribution

1. LEAEMPATY: A novel annotated empathy dataset (770 emails):
 - with paired examples (email1 & email2)
 - in German
 - first dataset for empathy in written leadership communication
2. Annotation scheme with specific behavioral indicators based on psychological theory and pre-existing scales (e.g. Sharma et al., 2021, Amjad et al., 2023)
3. Case Study demonstrating its utility for the development of computational empathy models via an empathy detection task

Data Collection: Online Experiment Jan 23



- Manager of a customer service department
- Previous manager was fired due to a mistake he made
- Orders from major customers have to be processed within 24 hours
- You were not available due to meetings, employee processed an order after consulting with colleagues but made a mistake → Lost customer
- Employee apologizes via e-mail, please write an answer.



Scan to access the data!

= 770 emails with an average length of 60 words

Empathy conceptualization

Affective empathy:

„Experiencing an affective state that is congruent with another person’s affective state.“



Cognitive empathy :

„Understanding another person’s internal state (i.e. her thoughts and affective state).“

- Existing scales designed for **in-person interactions** (e.g. Davis et al., 1980), **student feedback** (Wambsganss et al., 2021) **therapy context** (e.g Sharma et al. 2020)
- We develop a **new annotation scheme** for empathy annotations in text-based leadership communication.

Final Annotation Scheme for Empathy

-1	cognitive empathy failure , e.g. lecturing, blaming , own interpretations	affective empathy failure , e.g. emotional coldness, invalidation of feelings
0	Lack of empathy (symbolic category)	
+1	cognitive empathy success , e.g. perspective-taking, understanding	affective empathy success , e.g. validation, praise, appreciation

Annotation Scheme Example

		01 - cognitive empathy	
		working definition: detecting, recognizing and understanding others' cognitive and emotional states, meaning their thoughts, motifs and feelings.	
		Emphasis is on the observer <i>taking in the target's</i> mental state.	
points		definition & indicators	examples
a. failure	-1	01.a cognitive empathy failure definition: The observer does actively not put themselves in the target's position and instead offers their subjective interpretation of the situation as a fact. indicators: <ul style="list-style-type: none"> questioning/disbelief/doubt, disagreement/denial, offering opinion as fact, lecturing/preaching, blaming 	<input type="checkbox"/> questioning/disbelief/doubt: "it can hardly have been that difficult" denial/disagreement: "The statement that I was not available is completely lacking in any basis." offering opinion as fact: "You would have had a chance to contact me via the company chat and my secretary could have notified me." lecturing: "you should have coordinated this and found some way to do it despite everything." blaming: "This is your fault."

Annotation Process

1. Established universal coding rules
2. First round of coding in MAXQDA (Coder 1 and 2, first 25%)
3. Computation of Krippendorff's α : Acceptable for success categories, insufficient for failure
4. Discussion of inconsistencies and refinement of coding rules
5. Second round of coding in MAXQDA (Coder 1 and 2, next 25 %)
6. Krippendorff's α for the inter-annotator agreement was calculated again – revealing very good results (> 0.80)
7. The first annotator proceeded to code the remaining parts

Annotation Example

Example containing success and failure (ID 366, Email 1)

Hello Mr. Thiele,

I am aware of the difficulty of the decision you had to make and appreciate that you asked the team. However, this in no way excuses the mistake you made in selecting the devices.

I will inform you of any further decisions in this regard in a timely manner and until then (...).

Yours Sincerely

Example that only contains failure (ID 180, Email 1)

I was at least reachable by email and I expect you to take responsibility.

Example that only contains success (ID 487, Email 2)

Hello Mr. Thiele, thank you very much for your prompt and detailed response. I assure you that you don't have to feel bad about your mistake, because those who work make mistakes. I stand behind you in this matter and think it's time to make it clear to our clients that mistakes happen. We are very sorry for these mistakes, but it is all too human. Of course, we will compensate for the financial loss of our client. You have done what was necessary in your and my eyes, and for that, I thank you very much.

Key cognitive empathy: success failure affective empathy: success failure

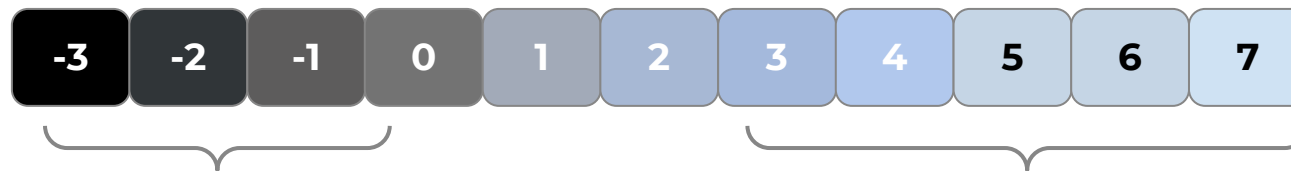
Case Study: Empathy Detection

Three setups:

1. Binary classification

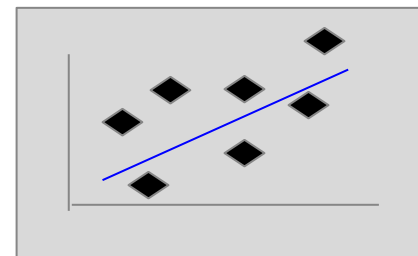


2. Multiclass classification



Few examples of these scores!

3. Regression to multiclass classification



Case Study: Models and Results

Experimented in 10-fold evaluation setup with a number of feature-based machine learning models and fine-tuning bert-base-german-cased.

- 1. Binary classification:** SVM with n-grams and LIWC-22 German features (lexicon of psycholinguistic categories of words)
- 2. Multiclass classification and regression to classification:** Fine-tuned bert-base-german-cased.

Task	Best model	Majority class	Prec.	Rec.	F1
BIN	SVM	57.1	81.7	81.8	81.7
MULTI	BERT	28.1	45.2	47.1	45.7
REGCLS	BERT	28.1	50.1	50.4	49.9

Table 4: Results for models with the best performance by weighted F1 score on each task.

Statistically significant performance over a majority class baseline

Future experiments could include Transfer Learning methods

Conclusions and Further Research

- Theory-based annotation scheme and LEAEMPATY Dataset
- Valuable resources for scholars aiming to
 - Understand the behavioral dimension of empathy
 - Compare aspects across domains and languages
- Detection tasks:
 - The models demonstrate they are learning signals of these empathy scores in preliminary benchmarks
 - Future work could explore more sophisticated methods to improve these results

Thank you!

Didem.Sedefoglu@uni-kassel.de
alahnala@uni-bonn.de