Universität Stuttgart
Institut für
Maschinelle Sprachverarbeitung

On the Complementarity of Images and Text for the Expression of Emotions in Social Media

Anna Khlyzova, Carina Silberer, and Roman Klinger

(1) Contribution

Corpus of Reddit posts
with text and image

Annotated via crowdsourcing for categories:

- Stimulus, Emotion
- Text–Image Relation

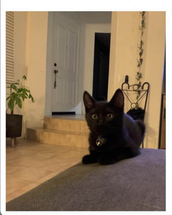
(2) Annotation Process

Please answer 3 questions about the post.

View the instructions for detailed instructions and examples BEFORE answering the questions. Please return the task if you dont know the answer. DO NOT answer the questions randomly as random answers can be recognized and will not be approved. Thank you!

Post

Newest member of the family, Bear



What emotion did the author likely feel when writing this post?

- joy
- surprise
- anticipation
- trust
- anger
- disgust
- fear
- sadness

What is the relation between the image and the text regarding emotion communication?

- The image is **complementary** to the text (the image is necessary to understand the emotion).
- The image is **illustrative** to the text (helpful to understand the emotion but not necessary).
- The image and text have emotions pulling in **opposite** directions when taken separately.
- The image is **decorative** to the text (redundant).
- The emotion is communicated with the **image only** (the text is annotation for the image).

What is it in the image that triggers the emotion? (Please choose the option that fits best)

- person/people
- animal
- object
- food
- meme
- screenshot/text in image
- art/drawing
- advertisement
- event/situation
- place
- other

(3) Examples

My everyday joy is to see my adorable cat smiles. And I've just realized, my cat can "dance with music". Amazing!
joy/complementary/animal

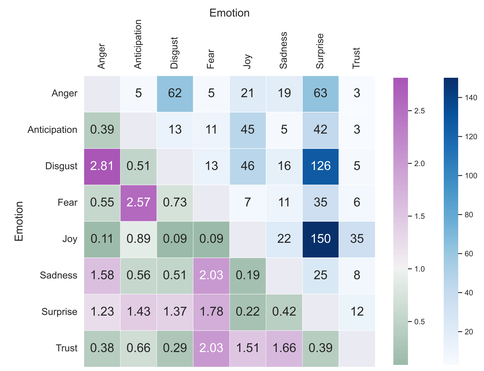
Don't move to Australia unless you can handle these bad boys
fear/complementary/animal

why didn't it fall
surprise/complementary/object

(4) Statistics

	Label	≥ 1	≥ 2	$= 3$	κ
Emo.	Yes	1,061	670	333	0.3
	No	1,047	710	319	0.3
Which emotion?	Anger	138	41	8	.26
	Anticipation	85	12	1	.11
	Disgust	268	127	57	.45
	Fear	64	15	5	.28
	Joy	585	444	329	.67
	Sadness	103	52	27	.56
	Surprise	435	221	84	.38
	Trust	54	6	1	.11
	Overall	*1732*	*918*	*512*	*.47*
Relation?	Complementary	1042	773	388	.02
	Decorative	124	6	0	.01
	Illustrative	476	152	4	.07
	Image only	142	27	0	.11
	Opposite	28	0	0	−.01
	Overall	*1812*	*958*	*392*	*.04*
Stimulus?	Advertisement	23	4	0	.14
	Animal	146	112	83	.79
	Art/drawing	157	58	33	.46
	Event/situation	132	27	2	.15
	Food	78	56	36	.74
	Meme	129	58	8	.34
	Object	211	102	51	.50
	Person	260	168	91	.61
	Place	46	12	5	.34
	Screenshot	528	351	195	.53
	Overall	*1710*	*948*	*504*	*.53*

(5) Corpus Analysis



(6) Modelling

- Early-fusion (tokens/preprocessed image)
- Late-fusion (output propabilities for each modality from RoBERTa and ResNet50)
- Model-based (input: penultimate layer of RoBERTa and ResNet50)

	Model	Emo.	Rel.	Stim.
uni-modal	Majority Baseline	.22	.56	.21
	Text	**.53**	**.77**	.45
	Image	.41	.67	.59

	Model	Emo.	Rel.	Stim.
multi-modal	Early fusion	.40	.72	.33
	Late fusion	.47	.72	.41
	Model-based fusion	**.53**	.76	**.63**

Contact: roman.klinger@ims.uni-stuttgart.de

Gefördert durch
DFG Deutsche Forschungsgemeinschaft