Automatically Generated Affective Norms of Abstractness, Arousal, Imageability and Valence for 350 000 German Lemmas

Maximilian Köper and Sabine Schulte im Walde
Institut für Maschinelle Sprachverarbeitung
Universität Stuttgart, Germany
{maximilian.koeper,schulte}@ims.uni-stuttgart.de

Affective Norms

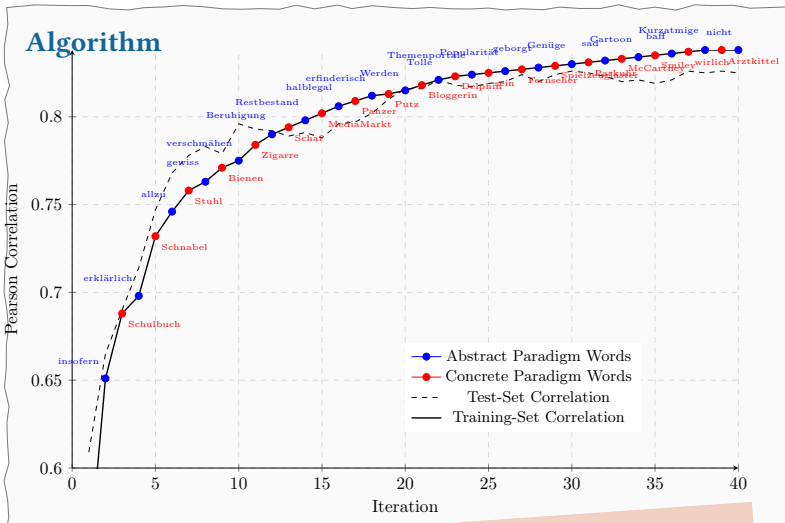
- **Abstract** words refer to things that we cannot perceive directly with our senses (*idea*, *politics*, ...), **concrete** words on the other hand refer to things that we can perceive (*image*, *scent*, ...,)
- A large subset of concrete words have a high **imageability**, these are words that refers to things that we can actually see
- **Valence** determines the pleasantness of a word (*gift* vs. *punishment*)
- **Arousal** describes the intensity of emotion provoked by a stimulus (*alert* vs. *calm*)

Training Data

Source	Words	Abs.	Ar.	Val.	Img.
Võ et al. (2009)	2902	✗	✓	✓	✓
Lahl et al. (2009)	2654	✓	✓	✓	✗
Kanske and Kotz (2010)	1000	✓	✓	✓	✗
Schmidtke et al. (2014)	1000	✗	✗	✗	✓
MRC∩Brys (*EN* → *GER*)	3266	✓	✗	✗	✗
# Unique words		5237	4848	4848	2901

- For every rating type we divided the number of unique words randomly into two sets: training data (90%) and test data (remaining 10%)

Algorithm



- We use the algorithm from Turney and Littman (2003)
- Idea: the degree of abstractness (or arousal, valence, imageability) can be expressed by comparing a given word w_i with a list of positive and a list of negative paradigm words
- Wordrepresentations are obtained via *word2vec* (skip gram)

$$R(w_i) = \sum_{p_j \in positive} sim(w_i, p_j) - \sum_{n_j \in negative} sim(w_i, n_j)$$

- The algorithm begins with an empty set of paradigm words and adds one word at a time to the paradigm list, alternating between positive paradigm words and negative paradigm words. At each step, we add the paradigm word that results in the highest Pearson correlation with the ratings of the training data

Turney and Littman (2003).
Measuring praise and criticism: Inference of semantic orientation from association
In *Journal ACM Transactions on Information Systems (TOIS)*

Final Correlation

	Abs	Ar	Val	Img
Training Correlation	0.838	0.796	0.827	0.832
Test Correlation	0.825	0.784	0.798	0.789

Data available!! www.ims.uni-stuttgart.de/data/affective_norms/

Examples

↑↓	Adj+Adv	*English*	Rating	Verb	*English*	Rating	NN	*English*	Rating
				Abstractness-Concreteness					
↑	aufgeblättert	exfoliated	8.44	entlangrutschen	slip along	7.97	Uniformtasche	uniform bag	10.00
↑	beinlang	leg-length	8.32	bekleckseln	blot	7.92	Vampirgebiss	vampire ivories	9.61
↓	paradox	paradox	0.63	negieren	negate	0.77	Selbstläuterung	self purification	0.52
↓	rechtfertigbar	justifiable	0.36	innewohnen	inhere	0.64	Willenlosigkeit	abulia	0.66
				Arousal					
↑	bestialisch	brutish	9.33	vergewaltigen	rape	9.85	Bandenrivalität	gang rivalry	10.0
↑	gewalttätig	violent	9.25	umbringen	kill	9.32	Blutbad	bloodbath	9.83
↓	satzweise	blockwise	0.81	flechten	weave	1.35	Wortfamilie	word family	0.48
↓	ausgerollt	rolled out	0.71	einfüllen	pour in	1.22	Holzdeckel	wood cover	0.00
				Imageability					
↑	neonerleuchtet	neon-lighted	9.83	weiterschnüffeln	continue sniffing	8.99	Granatangriff	grenade attack	10.0
↑	schwarzverhüllt	black-cloaked	9.61	emporzüngeln	upwardslicking	8.71	Polizeijeep	police Jeep	0.87
↓	gewiss	certain	0.58	auffassen	understand	1.08	Zielstellung	objective	0.59
↓	unstrittig	indisputable	0.41	elaborieren	elaborate	1.08	Sinnfreiheit	mindless	0.46
				Valence					
↑	wundervoll	wonderful	9.69	beschenken	endow	8.35	fitnessangebot	Fitness offer	10.0
↑	wunderbar	marvellous	9.69	genießen	enjoy	8.28	Frühlingsrezept	spring recipe	9.49
↓	katastrophenmässig	disastrous	0.39	zermürben	demoralize	0.60	Falschdiagnose	misdiagnosis	0.06
↓	ausblutend	bleeding to death	0.37	frikassieren	fricassee	0.57	Essensentzug	food deprivation	0.00