Supersense Embeddings: A Unified Model for Supersense Interpretation, Prediction, and Utilization

Lucie Flekova and Iryna Gurevych

Saarland University
Seminar: Embeddings for NLP and IR
Lecturer: Cristina España i Bonet

Aqsa Nazir & Egla Hajdini
July 10th, 2019

Table of content

Introduction:

- ➔ Supersenses are :
  - ◆ Coarse grained word labels
  - ◆ Based on WordNet's (Fellbaum,1998) lexicographer files
  - ◆ Contain 26 labels for nouns
  - ◆ Contain 15 labels for verbs

- ➔ Supersense labels have improved:
  - ◆ Dependency parsing (Agirre et al., 2011)
  - ◆ Named entity recognition (Marrero et al., 2009; Rud et al., 2011)
  - ◆ Non-factoid question answering (Surdeanu et al., 2011)
  - ◆ Question generation (Heilman, 2011)
  - ◆ Semantic role labeling (Laparra and Rigau, 2013)
  - ◆ Personality profiling (Flekova and Gurevych, 2015)
  - ◆ Semantic similarity (Severyn et al., 2013)
  - ◆ Metaphor detection (Tsvetkov et al., 2013)

Introduction (cont.):

- ➔ The first to provide a joint word and supersense-embedding model
- ➔ Provides an insight into the word and supersense positions in the vector space through similarity queries and visualizations
- ➔ A supersense tagging model which achieves competitive performance on recently published social media datasets was proposed
- ➔ How these predicted supersenses and their embeddings can be used in a range of text classification tasks

Table of content

Related Work: Semantically Enhanced Word Embeddings:

- ➔ Word sense or synset Embeddings usage is demonstrated in:
  - WSD (Word Sense Disambiguation)
  - Knowledge base Unification
  - Semantic Similarity
- ➔ Downstream document classification problems can be challenging

Related Work: Supersense Tagging

- ➔ Introduced by Ciaramita and Johnson (2003) for **nouns**
- ➔ Expanded for **verbs** (Ciaramita and Altun, 2006)
- ➔ Reimplemented by Heilman and widely used in applied tasks
- ➔ Johannsen et al. (2014) introduced a task of multiword supersense tagging on Twitter

Table of content

Building Supersense Embeddings:

- ➔ Mapping the Babel synsets to WordNet 3.0 synsets using the BabelNet API
- ➔ Mapping synsets to their corresponding WordNet's supersense categories.
- ➔ Using the Gensim implementation of Word2vec
- ➔ Applying the skip-gram model with negative sampling
  - ◆ continuous representations of words
  - ◆ supersense-disambiguated words
  - ◆ standalone supersenses

1	*About 10.9% of families were below the poverty line, including 13.6% of those under age 18.*
2	*About 10.9% of* N.GROUP *were below the* N.POSSESSION V.CHANGE *13.6% of those under* N.ATTRIBUTE *18.*
3	*About 10.9% of* FAMILIES_N.GROUP *were below the* POVERTY_LINE_N.POSSESSION INCLUDING_V.CHANGE *13.6% of those under* AGE_N.ATTRIBUTE *18.*

Table 1: Example of plain (1), generalized (2) and disambiguated (3) Wikipedia

Table of content

Qualitative Analysis: Verb Supersense

➔ Shows the most similar word vectors to each of the verb supersense vectors

➔ If no POS, they hold both semantic and syntactic information

VERBS	
BODY	wearing, injured, worn, wear, wounded, bitten, soaked, healed, cuffed, dressed
CHANGE	changed, started, added, dramatically, expanded drastically, begun, altered, shifted, transformed
COGNITION	known, thought, consider, regarded, remembered attributed, considers, accepted, believed, read
COMMUNICATION	stated, said, argued, jokingly, called, noted, suggested, described, claimed, referred
COMPETITION	won, played, lost, beat, scored defeated, win, competed, winning, playing
CONSUMPTION	feed, fed, employed, based, hosted feeds, utilized, applied, provided, consumed
CONTACT	thrown, set, carried, opened, laid pulled, placed, cut, dragged, broken
CREATION	produced, written, created, designed, developed directed, built, published, penned, constructed
EMOTION	want, felt, loved, wanted, delighted disappointed, feel, like, saddened, thrilled
MOTION	brought, led, headed, returned, followed left, turned, sent, travelled, entered
PERCEPTION	seen, shown, revealed, appeared, appears shows, noticed, see, showing, presented
POSSESSION	received, obtained, awarded, acquired, provided donated, gained, bought, found, sold
SOCIAL	appointed, established, elected, joined, assisted led, succeeded, encouraged, initiated, organized
STATIVE	included, held, includes, featured, served, represented, referred, holds, continued, related
WEATHER	glow, emitted, ignited, flare, emitting smoke, fumes, sunlight, lit, darkened

Table 2: Top 10 most similar word embeddings for verb supersense vectors

Qualitative Analysis: Verb Supersense

➔ Verb supersenses using the t-distributed Stochastic Neighboring Embedding

➔ It is a technique designed to visualize structures in high-dimensional data

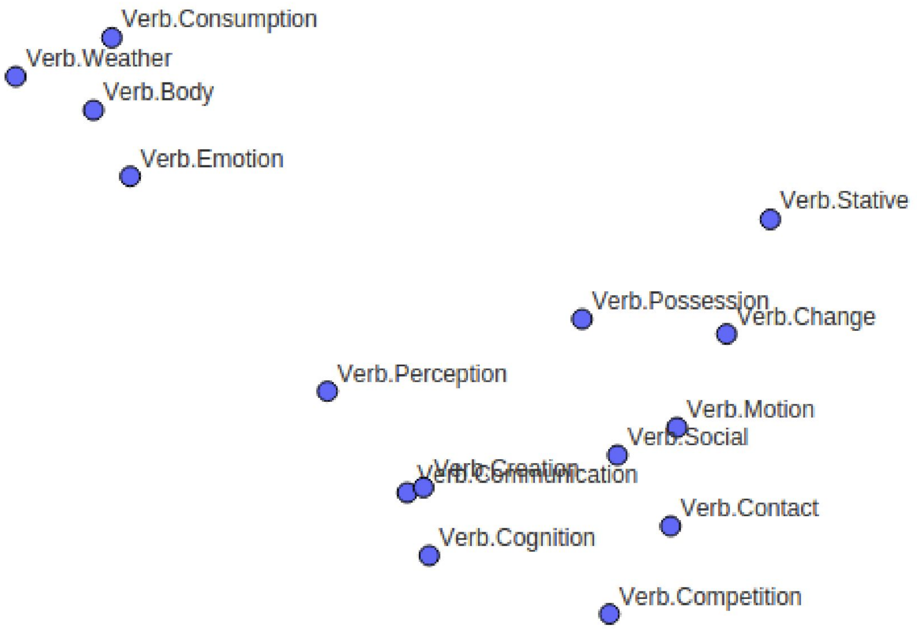

Figure 1: Verb supersense embeddings visualized in the vector space

Qualitative Analysis: Noun Supersense

- ➔ Most similar word embeddings for noun supersenses.
- ➔ FOOD, PLANT, TIME or PERSON are more plausible than those for abstract concepts such as ACT, ARTIFACT or COGNITION

NOUNS	
ACT	participation, activities, involvement, undertaken ongoing, conduct, efforts, large-scale, success
ANIMAL	peccaries, capybaras, frogs, echidnas, birds marmosets, rabits, hatchling, ciconiidae, species
ARTIFACT	wooden, two-floor, purpose-built, installed, wall fittings, turntable, racks, wrought-iron, ceramic, stone
ATTRIBUTE	height, strength, age, versatility, hardness power, fluidity, mastery, brilliance, inherent
BODY	abdomen, bone, femur, anterior, forearm femoral, skin, neck, muscles, thigh
COGNITION	ideas, concepts, empirical, philosophy, knowledge, epistemology, analysis, atomistic, principles
COMMUNICATION	written, excerpts, text, music, excerpted, translation, lyrics, subtitle, transcription, words
EVENT	sudden, death, occurred, event, catastrophic unexpected, accident, victory, final, race
FEELING	sadness, love, sorrow, frustration, disgust anger, affection, feelings, grief, fear
FOOD	cheese, butter, coffee, milk, yogurt dessert, meat, bread, vegetables, sauce
GROUP	members, school, phtheochroa, ypsolophidae pitcairnia, cryptanthus, group, division, schools
LOCATION	northern, southern, northeastern, area, south capital, town, west, region, city
MOTIVE	motivation, reasons, rationale, justification, motive justifications, motives, incentive, desire, why
OBJECT	river, valley, lake, hills, floodplain lakes, rivers, mountain, estuary, ocean
PERSON	greatgrandfather, son, nephew, son-in-law, father halfbrother, brother, who, mentor, fellow
PHENOMENON	wind, forces, self-focusing, radiation, ionizing result, intensity, gravitational, dissipation, energy
PLANT	fruit, fruits, magnifera, sativum, flowers caesalpinia, shrubs, trifoliate, vines, berries
POSSESSION	property, payment, money, payments, taxes tax, cash, fund, pay, $100
PROCESS	growth, decomposition, oxidative, mechanism rapid, reaction, hydrolysis, inhibition, development
QUANTITY	miles, square, meters, kilometer, cubic, ton, number, megabits, volume, kilowatthours
RELATION	southeast, southwest, northeast, northwest, east portion, link, correlation, south, west
SHAPE	semicircles, right-angled, concave, parabola, ellipse, angle, circumcircle, semicircle, lines
STATE	chronic, condition, debilitating, problems, health worsening, illness, illnesses, exacerbation, disease
SUBSTANCE	magnesium, zinc, silica, manganese, sulfur oxide, sulphate, phosphate, salts, phosphorus
TIME	september, december, november, july, april january, august, february, year, days
TOPS	time, group, event, person, groups

Table 3: Top 10 most similar word embeddings for noun supersense vectors

Qualitative Analysis: Noun Supersense

- ➔ Most similar word embeddings for noun supersenses.
- ➔ FOOD, PLANT, TIME or PERSON are more plausible than those for abstract concepts such as ACT, ARTIFACT or COGNITION

Figure 2: Noun supersense embeddings

Qualitative Analysis: Word Analogy & Word Similarity

- ➔ Assessing the changes between word embeddings WORDS and jointly with the supersenses-enriched SUPER
- ➔ Two evaluation tasks
- ➔ Noun supersense distinctions show the tendency to improve
- ➔ While syntax-related information is pushed to the background.

Group/Vectors:	WORDS	SUPER
Capitals - common	91.1	94.7±0.99
Capitals - world	87.6	89.5±0.69
City in state	65.2	65.7±1.03
Nationality to state	94.5	95.2±0.58
Family relations	93.0	94.4±1.28
Opposites	56.7	54.6±3.21
Plurals	89.4	86.4±1.08
Comparatives	90.6	90.4±0.85
Superlatives	79.4	79.6±1.83
Adjective to adverb	20.2	22.2±1.53
Present to participle	64.2	64.6±1.57
Present to past	60.0	59.2±1.30
3rd person verbs	84.3	82.1±1.44
Total	75.0	76.0±0.28

Table 4: Accuracy and standard error on analogy tasks

Qualitative Analysis: Word Analogy and Word Similarity

- ➔ The performance of WORDS and SUPER was compared on the following word similarity datasets:
  - WordSim353-Similarity (353-S)
  - WordSim353-Relatedness (353-R)
  - MEN dataset
  - RG-65 dataset (Rubenstein and Goodenough, 1965)
  - MC-30 (Miller and Charles, 1991)
- ➔ Better performance of vectors trained jointly with supersenses

Data:	MEN	353-S	353-R	RG-65	MC-30
WORDS	73.18	76.93	62.11	79.13	79.49
SUPER	74.26	78.63	61.22	79.75	80.94

Table 5: Performance of vectors (Spearman's ρ) on five similarity datasets.

Table of content

Building a Supersense Tagger

- A **window-based** approach with a **multi-channel multi-layer perceptron** model
- Extracting seven feature vectors for each word
  - 300 dimensional word embeddings
  - 41 cosine similarities of the word to each standalone supersense embedding
  - 41 cosine similarities of the word to each of its word_SUPERSENSE embeddings
  - Fixed vector of frequencies of each supersense in Wikipedia
  - Frequency of each word_SUPERSENSE in Wikipedia
  - Part-of-speech information
  - Casing information as a 3D unit vector

Building a Supersense Tagger

System/Data:	Tw-R-dev	Tw-R-eval	Tw-J-eval
Twitter-trained systems			
Searn (Johannsen et al., 2014)	67.72	57.14	42.42
HMM (Johannsen et al., 2014)	60.66	51.40	41.60
Ours Twitter (all features)	61.12	57.16	41.97
Ours Twitter no casing	61.06	56.20	41.13
Ours Twitter no similarities	63.47	56.78	39.44
Ours Twitter no frequencies	61.10	57.32	39.02
Ours Twitter no part-of-speech	57.08	54.45	36.50
Ours Twitter no word embed.	57.57	53.43	34.91

Table 6: Weighted F-score performance on supersense prediction

Table of content

Using Supersense Embeddings in Document Classification Task

- Three channels of word embeddings:
    - 300-dimensional word embeddings
    - 41-dimensional vectors capturing the cosine similarity of the word to each supersense embedding
    - vector of relative frequencies of the word occurring together with its supersense
- In parallel a processed document text of predicted supersenses

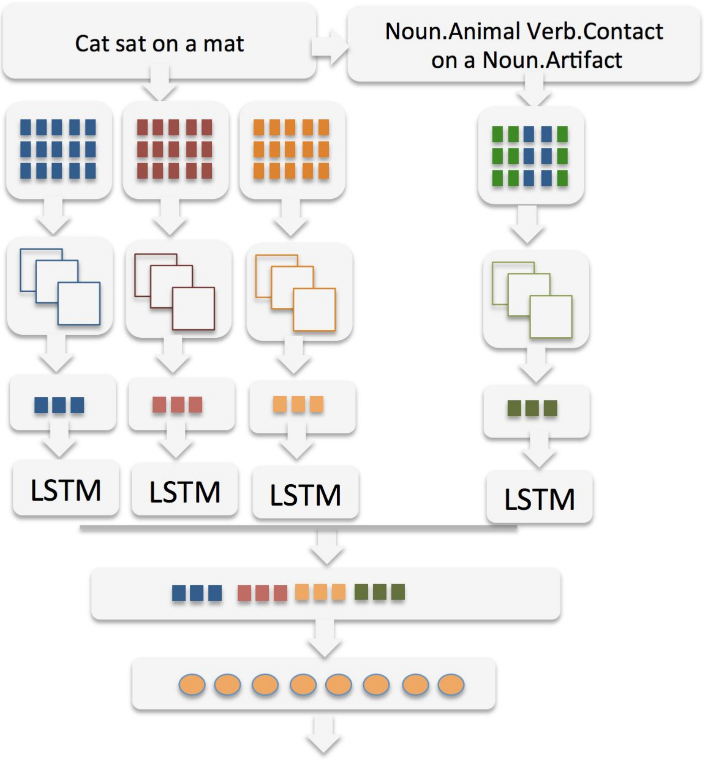

Figure 3: Network architecture

Sentiment Polarity Classification

- Binary sentence classification
- The Movie Review dataset
- Supersenses help to generalize over rare terms

➕ GROUP, LOCATION, TIME and PERSON

➖ PERCEPTION, SOCIAL and COMMUNICATION

System	Accuracy
Socher et al. (2011)	77.7
Socher et al. (2012)	79.0
Wang and Manning (2013)	79.1
Dong et al. (2015)	79.5
Kim (2014)	81.5
WORDS	79.4
SUPER	$\boxed{\mathbf{81.7}\pm 0.37}$

Table 7: 10-fold CV accuracy and standard error for the sentiment classification task

Sentiment Polarity Classification

Positive reviews	
Text	Supersenses
beating the austin powers film at their own game , this blaxploitation spoof downplays the raunch in favor of gags that rely on the strength of their own cleverness as oppose to the extent of their outrageousness .	verbstative the nounlocation nouncognition nounartifact at their own nouncommunication , this nounact nouncommunication verbstative the nouncognition in nouncommunication of that verbcognition on the nouncognition of their own nouncognition as verbcommunication to the nounevent of their nounattribute .
there is problem with this film that even 3 oscar winner ca n't overcome , but it 's a nice girl-buddy movie once it get rock-n-rolling .	there verbstative nouncognition with this nouncommunication that even 3 nounevent nounperson ca n't verbemotion , but it verbstative a nice girl-buddy nouncommunication once it verbstative rock-n-rolling
godard 's ode to tackle life 's wonderment is a rambling and incoherent manifesto about the vagueness of topical excess . in praise of love remain a ponderous and pretentious endeavor that 's unfocused and tediously exasperating .	nounperson nouncommunication to verbstative nouncognition 's nouncognition verbstative a rambling and incoherent nouncommunication about the nounattribute of topical excess . in nouncognition of nouncognition verbstative a ponderous and pretentious nounact that verbstative unfocused and tediously exasperating
Negative reviews	
Text	Supersenses
the action scene has all the suspense of a 20-car pileup , while the plot hole is big enough for a train car to drive through – if kaos have n't blow them all up .	the nounact nounlocation verbstative all the nouncognition of a 20-car nouncognition , while the nounlocation verbstative big enough for a nounartifact nounartifact to verbmotion through – if nounperson have n't verbcommunication them all up .
the scriptwriter is no less a menace to society than the film 's character .	the nounperson verbstative no less nounstate to noungroup than the nouncommunication nounperson .
a very slow , uneventful ride around a pretty tattered old carousel .	a very slow , uneventful nounact around a pretty tattered old nounartifact .
the milieu is wholly unconvincing . . . and the histrionics reach a truly annoying pitch .	the nouncognition verbstative wholly unconvincing and the nouncommunication verbstative a truly annoying nounattribute .

Table 8: Example of documents classified incorrectly with word embeddings and correctly with word and supersense embeddings

Subjectivity Classification

- Supersenses are a natural candidate for subjectivity prediction
- Nouns and Verbs in the subjective and objective sentences come from different semantic classes
- VERB.FEELING vs. VERB.COGNITION
- PERCEPTION, COMMUNICATION, ATTRIBUTE
- SOCIAL, PERSON, POSSESSION

System	Accuracy
SVM (Pang and Lee, 2004)	90.0
NB (Pang and Lee, 2004)	92.0
CNN (Kim, 2014)	93.4
F-Dropout (Wang and Manning, 2013)	93.6
MV-CNN (Zhang et al., 2016)	93.9
WORDS	92.1
SUPER	$\boxed{\mathbf{93.9}\pm0.26}$

Table 9: Binary classification on the subjectivity dataset

Metaphor Identification

- Supersenses hold the information of coarse semantic concepts
- Only a subset of the architecture is used
  - word embeddings
  - similarity vectors
  - supersense frequency vectors

System	F1-score on test set
(Gershman et al., 2014)	85
WORDS	81.91 ± 2.81
SUPER	$\boxed{\mathbf{87.23} \pm 2.36}$

Table 10: F1-score and a standard error on a provided test set for the adjective-noun metaphor prediction task

Table of content

Conclusions:

- ➕ A novel joint embedding set of words and supersenses
- ➕ It provides a new insight into the word and supersense positions in the vector space
- ➕ A technique for learning supersense representations, using automatically-annotated corpora
- ➕ The supersense enrichment can be beneficial to a range of binary classification tasks
- — Complexity of the model, additional sources of knowledge

Questions?

References:

[1] L. Flekova, I. Gurevych “Supersense Embeddings: A Unified Model for Supersense Interpretation, Prediction, and Utilization” ACL 2016

https://github.com/UKPLab/ acl2016-supersense-embeddings

https://github.com/coastalcph/ supersense-data-twitter

https://www.cs.cornell.edu/people/ pabo/movie-review-data/

http://www.cs.cmu.edu/˜ytsvetko/ metaphor/datasets.zip