Probabilistic Soft Logic for Semantic Textual Similarity

Islam Beltagy, Katrin Erk, Raymond Mooney

The University of Texas at Austin

Abstract

-Using **probabilistic logic** for semantic representation, combining **Formal Semantics** and **Distributional Semantics**

-**PSL** is the probabilistic logic framework we use

-Evaluate on the **STS** task: judge sentence similarity

Task

Compute the **Semantic Similarity** score of the given sentences

Formal Semantics

(deep but brittle)

-represent semantics using **formal languages** like FOL.

-handles many complex semantic phenomena like embedded propositions, logical operators, and quantifiers

-unable to handle uncertain knowledge

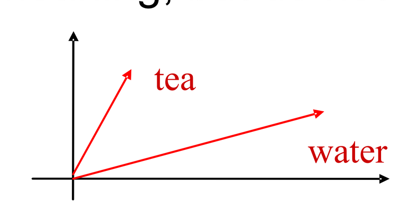
Distributional Semantics

(robust but shallow)

-represent words as vectors in high dimensional space

-capture the "graded" nature of linguistic meaning, but do not adequately capture logical structure

-similarity(water, tea) = cos(water, tea)



Probabilistic Logic

-Probabilistic Logic combines Formal Semantics and Distributional Semantics

-Probabilistic Logic offers a **deep and robust** semantic representation

Probabilistic Soft Logic (PSL)

-PSL is the probabilistic logic framework we use

-Atoms have **continuous truth values** in interval [0,1] (in contrast with boolean atoms in MLNs)

-**Efficient inference** (100 times faster than MLN in our experiments)

-Logical operators are replaced with Łukasiewicz logic:

$I(\neg \ell_1) = 1 - I(\ell_1)$

$I(\ell_1 \wedge \ell_2) = \max \{0, I(\ell_1) + I(\ell_2) - 1\}$

$I(\ell_1 \vee \ell_2) = \min \{1, I(\ell_1) + I(\ell_2)\}$

-Implication via distance to satisfaction ***d*** : $d(\ell_1 \rightarrow \ell_2) = \max \{0, I(\ell_1) - I(\ell_2)\}$

-PDF: $f(I) = \frac{1}{Z}\exp[-\sum_{r \in R} \lambda_r (d_r(I))^p]$

-MPE inference is a **linear program** when p = 1

Adapting PSL for STS

-PSL does not work out-of-the-box for STS

-Replace conjunction with average: $I(\ell_1 \wedge \dots \wedge \ell_n) = \text{avg}(I(\ell_1), \dots, I(\ell_n))$

-Inference remains a linear program

-Replace PSL grounding with a **heuristic grounding** that find the **n** best groundings (grounding limit)

System Architecture

"A man is driving a convertible."

"A guy owns a nice car."

Boxer

Semantic analysis tool maps sentences to logical form

S1: ∃x,y,z. man(x) ∧ drive(y) ∧ agent(y, x) ∧ convertible(z) ∧ patient(y, z)

S2: ∃x,y,z. guy(x) ∧ own(y) ∧ agent(y, x) ∧ nice(z) ∧ car(z) ∧ patient(y, z)

Vector Space

Distributional Rules Constructor

Uses distributional information to generate an **on-the-fly soft rule base**

∀x. man(x) → guy(x) | sim(man, guy)

∀x. drive(x) → own(x) | sim(drive, own)

∀x. man(x) → car(x) | sim(car, guy)

∀x. convertible(x) → nice(x) ∧ car(x) | sim(convertible, "nice car")

.....

Rules from Paraphrase collections like PPDB

Rule Base (RB)

A set of **weighted rules representing background knowledge**

PSL inference

Computes:

`P(S1|S2, RB)`

`P(S2|S1, RB)`

Result

Semantic Similarity score = `F( P(S1|S2, RB), P(S2|S1, RB) )`, `F` is learned

Heuristic Grounding

Evidence: I(man(M)) = 1, I(drive(D)) = 0.8 Constants: M, D

Rule: man(x) ∧ drive(y) ∧ agent(y, x) → result()

score(x=D) = 0 + 0 = 0	man(D) ∧ agent(D, D) ∧ drive(D)	score(y=D) = 0.8 + 0 = 0.8
	man(D) ∧ agent(M, D) ∧ drive(M)	
score(x=M) = 1 + 0 = 1	man(M) ∧ agent(M, M) ∧ drive(M)	score(y=M) = 0 + 0 = 0
	man(M) ∧ agent(D, M) ∧ drive(D)	

Best ground clause is when {x=M, y=D}: man(M) ∧ agent(D, M) ∧ drive(D)

Evaluation

Correlation

System	msr-vid	msr-par	SICK
vec-add	0.78	0.24	0.65
MLN	0.63	0.16	0.47
PSL	0.79	**0.53**	0.70
PSL+vec-add	**0.83**	0.49	**0.71**

Time

Dataset	PSL (Time)	MLN (Time / Timeouts)
msr-vid	8s	1m 31s / 9%
msr-par	30s	11m 49s / 97%
SICK	10s	4m 24s / 36%

Conclusion

1) Probabilistic logic in general is a promising semantic representation

2) PSL fits the STS task better than MLNs. It is faster and more accurate

References [PSL] S. Bach, M. Broecheler, L. Getoor, D. O'Leary. 2012. Scaling MPE Inference for Constrained Continuous Markov Random Fields with Consensus Optimization. NIPS 2012 [MLN] M. Richardson and P. Domingos. Markov logic networks. Machine Learning 2006 [BOXER] J. Bos. 2008. Wide-coverage semantic analysis with boxer. STEP 2008