

Vocabulary Expansion by Semantic Extraction of Medical Terms

Maria Skeppstedt¹, Magnus Ahltop², Aron Henriksson¹

¹*Department of Computer and Systems Sciences (DSV),
Stockholm University, Sweden*

²*Royal Institute of Technology, Sweden*

Aim

Automatically extract terms that belong to a specific semantic category.

Applications

Can be used for, e.g.:

- ▶ Semi-automatic terminology development
- ▶ Supporting named entity recognition

Chosen semantic categories

Terms denoting:

- ▶ Medical Finding
- ▶ Pharmaceutical Drug

Used resources

- ▶ A large corpus of **Swedish medical text**: Lakartidningen*
- ▶ Terms from the Swedish translation of MeSH belonging to the semantic categories:
 - ▶ **Disease or syndrome** and **Sign or symptom** (for Medical Finding)
 - ▶ **Pharmacologic substance** (for Pharmaceutical Drug)

* The Journal of the Swedish Medical Association, 21 447 900 tokens, 444 601 unique terms

Underlying idea: Distributional semantics

Harris (1954): Words with similar meanings tend to occur in similar contexts

The patient complained of itching *dermatitis*

The patient complained of itching *eczema*

Random indexing: A type of word space model

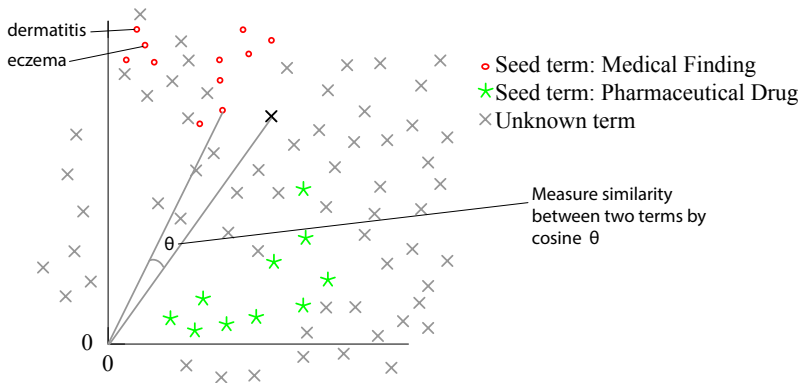
Word co-occurrence information was represented by a word space model.

A computationally light-weight version of the word space model was used, called **random indexing*** .

* See: Random Indexing of Text Samples for Latent Semantic Analysis, Kanerva et al. (2000), The Word-Space Model Using distributional analysis to represent syntagmatic and paradigmatic relations between words in high-dimensional vector spaces, Sahlgren (2006).

Word space model

A hypothetical word space model with a dimensionality of 2.



How the MeSH terms were used

Divide MeSH terms into 2 equally large sets:

- ▶ A set to use as seed terms
(91 Medical Findings, 91 Pharmaceutical Drugs)
- ▶ A set to use for evaluation,
representing unknown words we want to find
(90 Medical Findings, 90 Pharmaceutical Drugs)

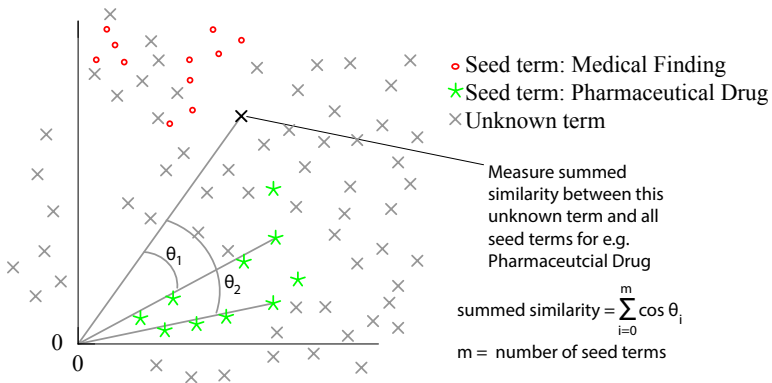
Extract terms close to seed terms,
evaluate recall against evaluation set.

Two methods for extracting terms close to seed terms

- ▶ Summed similarity to seed terms (CosAdd)
- ▶ Replace all occurrences of seed terms in the corpus with a common string (TermRep)

Summed similarity to seed terms (CosAdd)

- ▶ Rank all unknown terms according to summed similarity.
 - ▶ Measure recall for top n.



Replace all occurrences of seed terms in the corpus with a common string (TermRep)

The patient complained of itching *dermatitis*

The patient complained of itching *eczema*

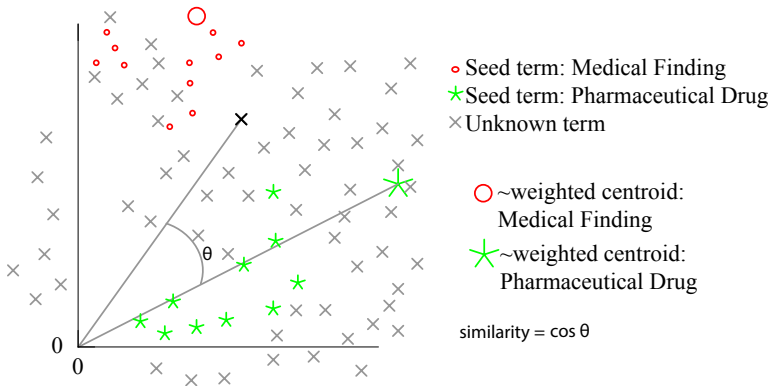
→

The patient complained of itching *FINDING*

The patient complained of itching *FINDING*

Replace all occurrences of seed terms in the corpus with a common string (TermRep)

- ▶ Rank all unknown terms according to similarity.
 - ▶ Measure recall for top n.



Setting parameters

- ▶ Dimensionality for the word space: 1000
- ▶ Cut-off for unknown terms and MeSH terms to be included: more than 50 occurrences
- ▶ Evaluated a number of window sizes...

Evaluation of window size

The patient in question had itching dermatitis on the back of the right hand.

Evaluated window sizes:

1+1

2+2

4+4

50+50

Evaluation data for window size

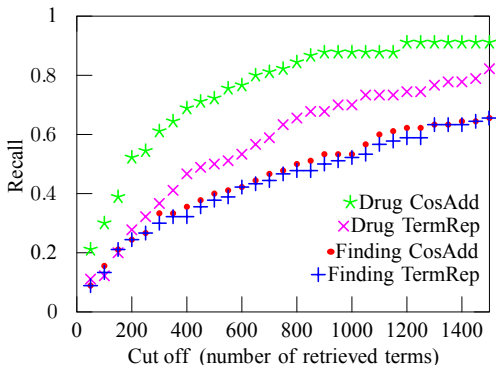
- ▶ 10-fold cross-validation on seed terms
 - ▶ Used 9/10 as seed terms
 - ▶ Measure recall for retrieving 1/10 held-out terms among top: 50, 100, 150 ... 1000

Results for evaluation of window size

No large differences

Window Size	1+1	2+2	4+4	50+50
	Medical Finding			
CosAdd	0.372	0.389	0.384	0.382
TermRep	0.357	0.368	0.361	0.360
	Pharmaceutical Drug			
CosAdd	0.567	0.516	0.502	0.501
TermRep	0.409	0.386	0.375	0.371

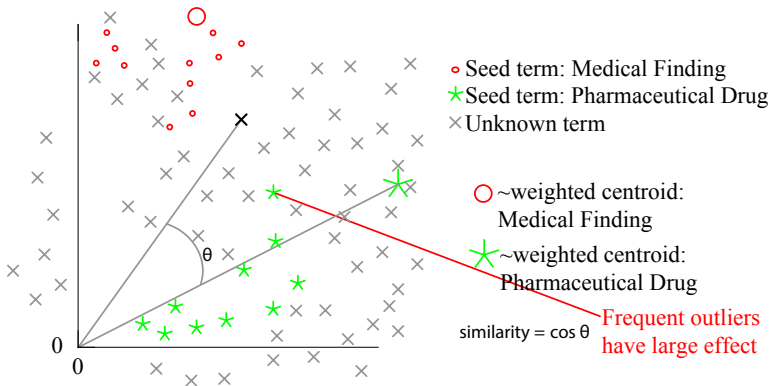
Results on evaluation term set



CosAdd: For identifying 90 unknown terms among top 1000:

- ▶ Medical Finding: Recall 0.53
- ▶ Pharmaceutical Drug: Recall 0.88

Why is CosAdd better?



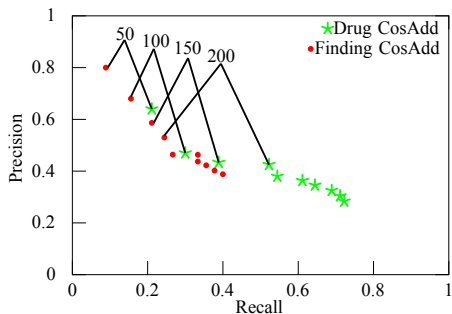
Why is recall not better for Medical Finding?

CosAdd: For identifying 90 unknown terms among top 1000:

- ▶ Medical Finding: Recall 0.53
- ▶ Pharmaceutical Drug: Recall 0.88

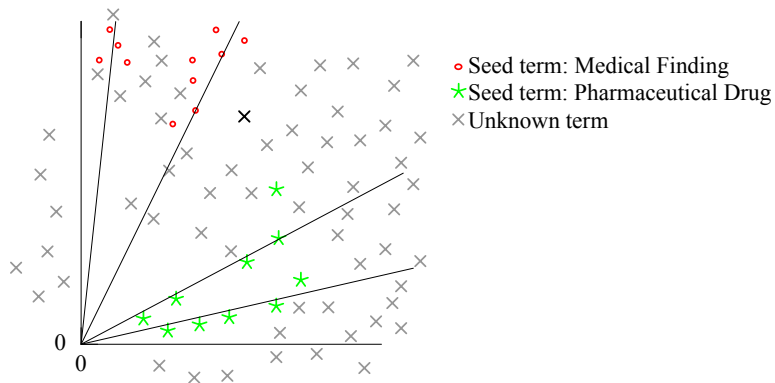
Why is recall not better for Medical Finding?

There are a lot of other medical findings, apart from the 90 that are searched for



Precision (partially based on manual classification) vs. recall (automatically measured against the reference standard), cut-off 50–500

Next step



Questions

- ▶ What method should the methods explored here be compared to?
- ▶ We tried the same method on Japanese patient blog texts, but failed completely.
Any suggestions why?

Thanks!