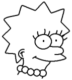
Predicting Cognitively Salient Modifiers of the Constitutive Parts of Concepts

Gerhard Kremer   Marco Baroni

CIMeC, University of Trento, Italy

July 15, 2010

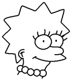
Describing a Concept...



- has 4 paws
- has a tail
- barks

Describing a Concept. . .

“Dog”



- has 4 paws
- has a tail
- barks

– vs. –

- has a heart
- can see

Topic and Focus

Feature Norms (e. g. McRae et al.'s)

Concept representations – used in simulations of cognitive tasks

Topic and Focus

Feature Norms (e. g. McRae et al.'s)

Concept representations – used in simulations of cognitive tasks

Efforts on extracting such descriptions

. . . using text corpora

(getting norms without experiments;

better models based on more data)

Topic and Focus

Feature Norms (e. g. McRae et al.'s)

Concept representations – used in simulations of cognitive tasks

Efforts on extracting such descriptions

. . . using text corpora

(getting norms without experiments;

better models based on more data)

(New) focus here

Composite part properties (adj modifier + noun) of concepts,

e. g. *rabbit*: *long* *ears*

Approach

- **Aim**:
  Extract cognitively salient modifiers
  for given concept–part pairs
- **Idea**:
  Create ranked list based on corpus frequencies
  and select 5 highest ranked modifiers
- **Resource**:
  WaCky web corpus
- **Evaluation** against feature production norms

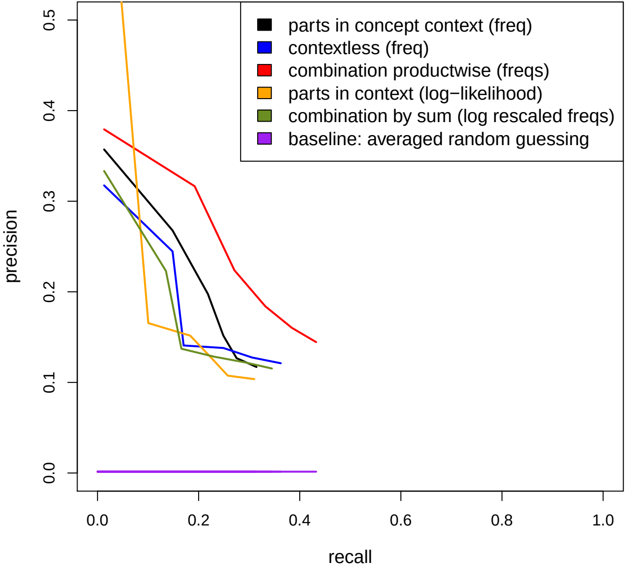
Rank List Methods

1. Modifier–Part pair frequencies ("contextless")
   [*Adj*]? [*Adj*]? [*Adj*]? [*Adj*]? [*Noun*]
2. Log-Likelihood ratios of frequencies
3. Frequencies of modifier–part pairs in concept context
   [*part*]? (20 *sent.*) [*concept*] (20 *sent.*) [*part*]?
4. Summed log-rescaled frequencies
5. Productwise combination of frequencies

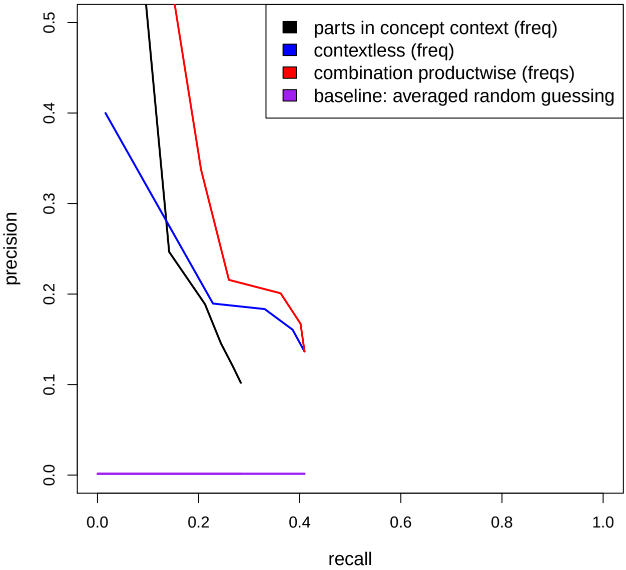
Example:
Concept “Bear” With Part “Fur”

	contextless		in concept context	
rank	freq	modifier	freq	modifier
1	507	thick	16	thick
2	209	dense	14	white
3	204	soft	11	small
4	185	black	11	soft
5	175	long	9	dense

Performance (GER)



Performance (ITA)



Plausibility Judgements (GER)

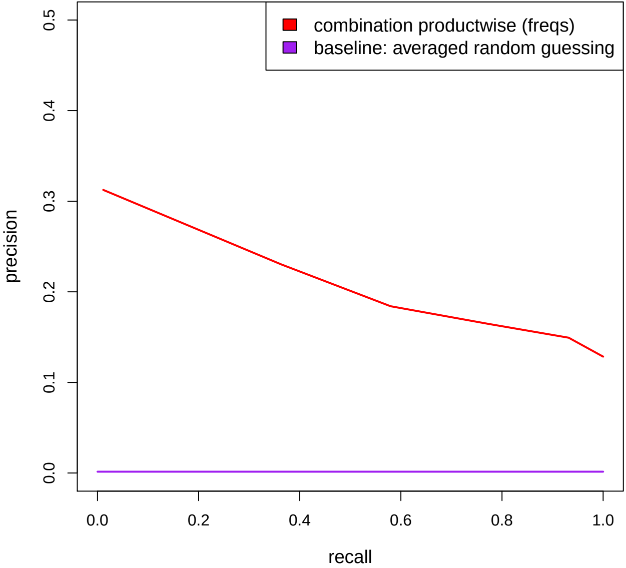
Setting

- Top 5 candidates of best method (productwise combination)
- "The *part* of a *concept* is *modifier*."
- Plausible/unlikely to be used in concept explanation?

Evaluation

... for those concept–modifier–part triples with acceptance $\geq 75\,\%$

Performance Based on Plausibility Ratings (GER)



Discussion

Automatic corpus-based extraction

... works best when combining in-context and contextless list

... performs similarly well across languages

... works comparably well based on both production and perception

Further Work

Extension

- Include numerals
- Decide if modifier necessary for specific part

Evaluation

Filter unlikely modifiers (more production data, judgements)

Next

- Salient parts (as preceding step)
- Extract other relation types

. . . thank you.

Selected Literature

McRae, K., Cree, G., Seidenberg, M., and McNorgan, C. (2005).
Semantic feature production norms for a large set of living and nonliving things.
Behavior Research Methods, 37(4):547–559.

Spence, D. and Owens, K. (1990).
Lexical co-occurrence and association strength.
Journal of Psycholinguistic Research, 19(5):317–330.