Distributional semantics from text and images

2011

Bruni, Tran & Baroni

Spinach – Two representations

- Linguistic representation – Statistical large corpus
  - plant, green, iron, popeye, muscles
- non-linguistic representations - perceptual information
  - our experience with it, its colours, smell, etc…
- Get both representation to communicate
  - Richer
  - More human like

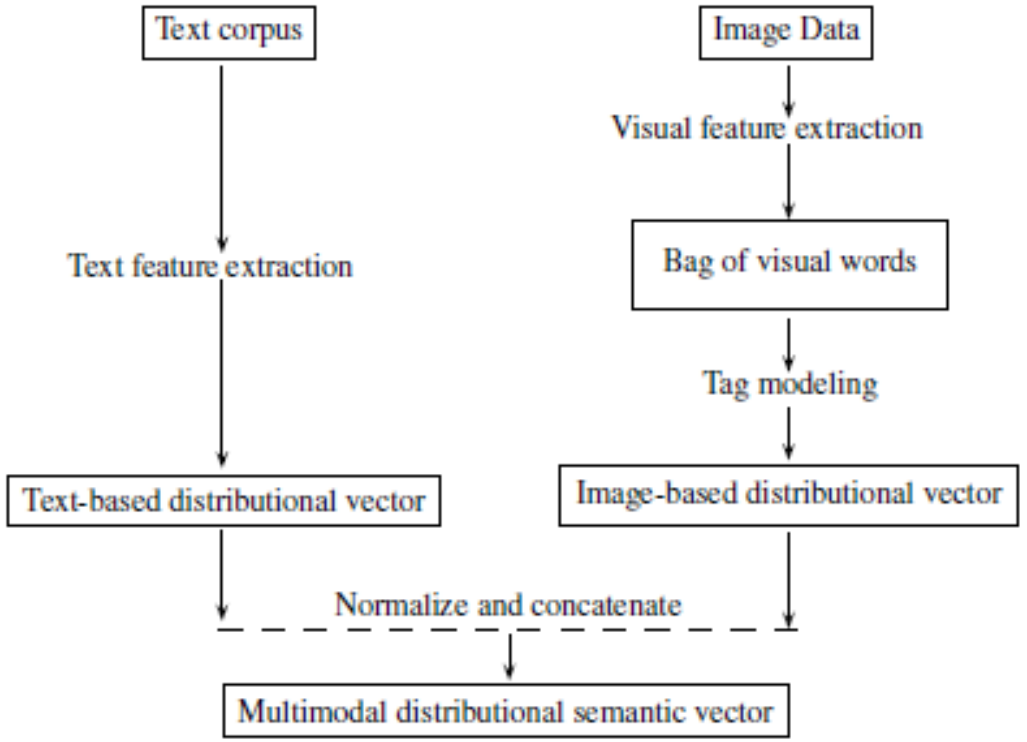
Images and text

- Perceptual information coming from images
  - NLP and computer vision
- multimodal distributional semantic model extracted from texts and images
  - Corpus of tagged images
  - build vectors recording the cooccurrences of words with image-based features
  - concatenate the image-based vector with a standard text-based distributional vector

overview

1. Introduction
2. overview
3. Implementation of model
    1. Overview
    2. Text
    3. Image
    4. combining
4. Results
5. conclusions

Proposed method - overview



Bag of visual words

- Similar to that of Bag of words
- Feature extraction -> quantize vectors -> histogram

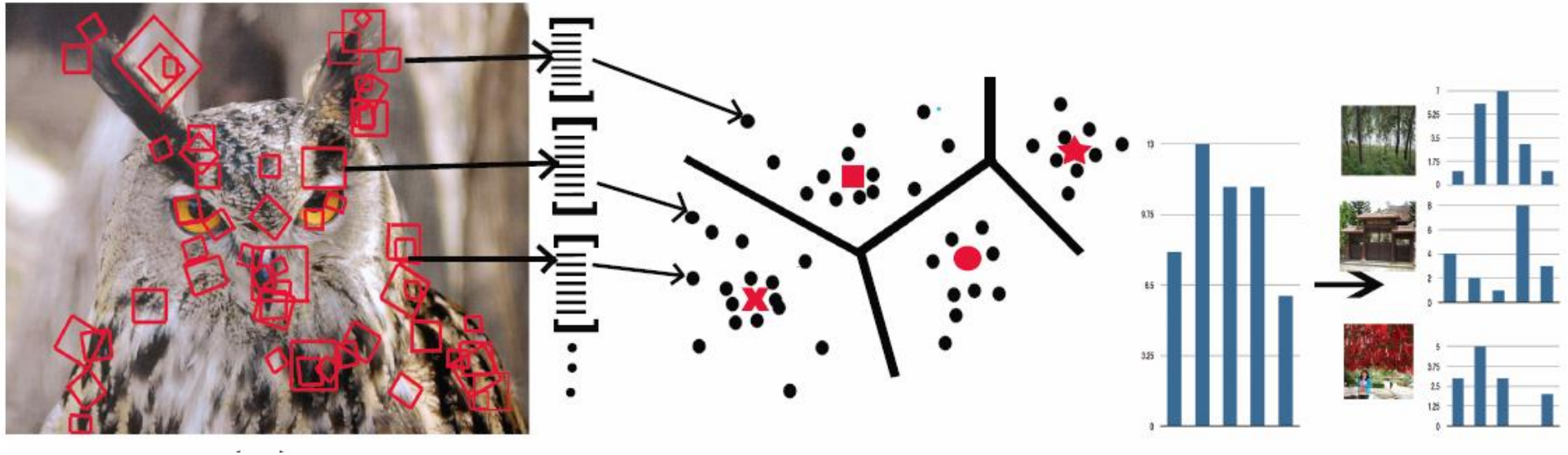


Image-based distributional model

- Tagged ESP-Game data set
- co-occurrence count : Sum visual word count for the tag
- Convert these score to an approximated log-likelihood score
- Matrix model
  - Each row corresponds to a tag vector
  - Summary of the distribution of visual words

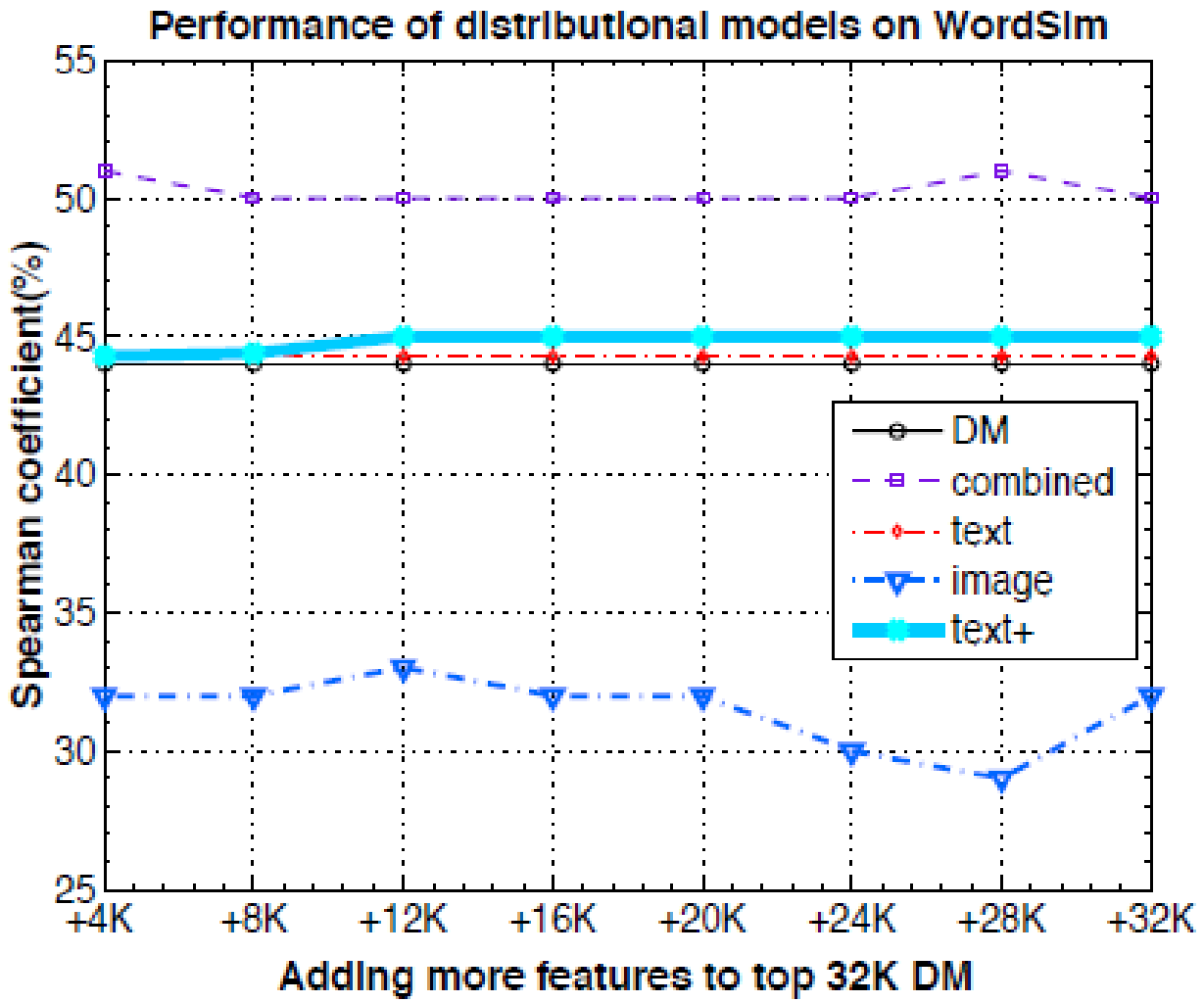
Text-based distributional vector

- TypeDM – Baroni and Lenci
  - 2.8 billion tokens , State of the art, designed to be used by others
- Collocates- labelled with the link that connect them to the target words
- Links - mixture of dependency parse information and lexico-syntactic patterns
  - Subject_kill, with_gun or as_sharp_as
- score
  - different surface realizations
- fat and the feature of_animal
  - fat of the animal, the fat of the animal, fats of animal.

Model integration

- Normalize : the image and text vectors to length 1
- Only use the top n dimensions for text vectors
- Concatenate the two vectors
  - Very simple
- Text+ - increase the dimensions to be equal to that of combined

Results – WordSim



Examples

combined	*text+*
tennis/racket	physics/proton
planet/sun	championship/tournament
closet/clothes	profit/loss
king/rook	registration/arrangement
cell/phone	mile/kilometer

Results - Concept categorization

model	*AP*	*Battig*
DM	81	96
text	79	83
text+	80	86
image	25	36
combined	78	96

Result's - BLESS

- 174 concepts each paired with a set of words that instantiate the following 6 relations: hypernym, coordination, meronym, typical attribute and typical event
- Compare DM and pure image model
- Image best at capturing differences attributes/events and adjectives/verbs
- Bias towards nouns
- Image is best at colours and shapes
- Syntax best at systemic or functional characteristics such as powerful or elegant
- Combined best of both worlds.

Conclusion

- Adding images is not damaging
- Images are better at concrete ideas where as text is better at abstract ideas
- Basic
- Further work is planned to expand upon these ideas