Composition in Distributional Models of Semantics

Jeff Mitchell, Mirella Lapata, 2010

Reporter: Dongyao Hu (胡东瑶)

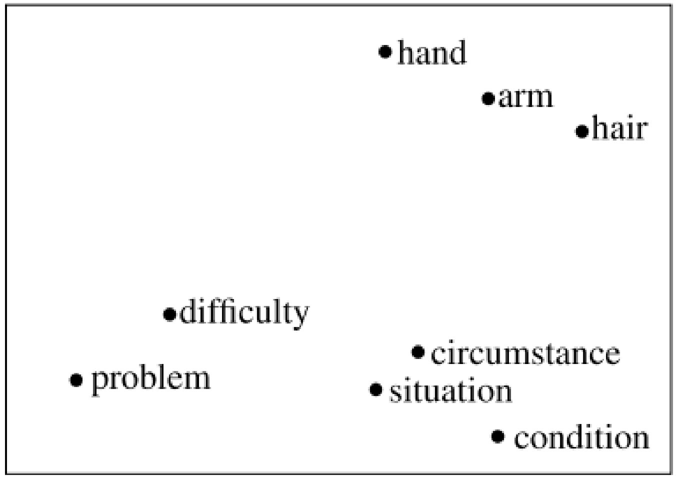
Semantic representation for words

- A topic for great debate in cognitive science history.
- A lot of use and influences.
- Fall under **three broad families:**
  - **Semantic networks:** represent concepts as nodes in a graph
  - **Feature-based models:** word meaning can be described in terms of feature lists
  - **Semantic spaces:** word meaning can be learned from the linguistic environment

Semantic representation for words

Semantic spaces

- In a semantic space **words are represented as points**, and **proximity indicates semantic association**.
  - **Hyperspace Analog to Language model (HAL)**
  - **Latent Semantic Analysis (LSA): word–document co-occurrence matrix** from a large document collection
  - etc.



Isolation vs. Composition

Words vs. Phrases etc.

- These models are typically directed at representing words in isolation
- Linguistic structures are **compositional**
  - Morphemes (词素)
  - Words
  - Phrases
  - Sentences
- Semantic space models **can naturally measure similarity** but are **not compositional**.

Semantic representation for phrases etc.

- **Logic-based view** 本文借鉴其方法
  - The meaning of a phrase or sentence is its **truth conditions (真值条件)** which are expressed in terms of truth relative to a model.
- **Connectionism** 本文借鉴其方法
  - Words **shouldn’t** be represented as **discrete symbols** that enter into symbolic expressions.
  - **Should** be represented as **vectors** of activation **distributed over many units (i.e. neurons.)**
  - The key is to **bind** one vector to another, e.g. by **tensor product** (Smolensky, 1990).
- **Semantic spaces** 本文采用这种词向量表示
  - Words occurring within similar contexts are semantically similar.
  - 和 Connectionist 的方法类似，但是向量不是 binary 也不是 randomly distributed，而是对应通过上下文得到的 **co-occurrence counts**。

Semantic representation for phrases etc.

- **Logic-based view**
  - Montague grammar
- **Connectionism**
  - Tensor products
  - Recursive distributed representations
  - Spatter codes
  - Holographic (全息的) reduced representations
  - Convolution
- **Semantic spaces**
  - Vector averaging (discard word order and syntactic structure)
  - Tensor products

Composition

- **Definition** of compositionality
  - Meaning of a whole = f(part1, part2)
  - Meaning of a whole = f(part1, part2, syntact)
  - Meaning of a whole = f(part1, part2, syntact, knowledge)

- **Difficulty** in defining compositionality
  - The meaning of the whole is constructed from its parts.
  - The meaning of the parts is derived from the whole.
  - Compositionality is a matter of **degree**
    - Fully. E.g., black hair
    - Partly. E.g., take advantage
    - None. E.g., kick the bucket (去世) 固定搭配

Compositional models

Goal & Question

- **Goal**
  - Construct vector representations for phrases and sentences.
- **Question** 质疑
  - whether representations in **a fixed space** are **flexible enough** to **cover the full expressivity** of language.
- **Answer**
  - 限制词组、句子为定长其实也是在限制模型结构的复杂度。
  - 使得计算可行，不必为每种结构设计不同的处理方法。
  - 实践中，构建向量比构建矩阵容易。
  - 本论文的方法只能用于单一维度空间，无法用于多个不同维度的空间。

Compositional models

Mathematical types of function

- **Linear composition**
  - p=Au+Bv, where A and B are matrices.
    - p=u+v, **additive**.
- **Tensor production**
  - p=Cuv, C is a tensor of rank 3.
  - p=u⊙v, element-wise prod (Hadamard, Schur, entrywise, pointwise prod).
    - pi=ui*vi, also called **multiplicative**.
  - p=u⊗v, **tensor prod** (corss, vector , dyadic, outer prod).
  - P=conv(uv), **circular convolution**.
    - u 和 v 的大小只能影响 p 的大小，不能影响 p 的方向。
    - 传统的加法模型中，u 和 v 的大小既影响 p 的大小，又影响 p 的方向。
    - Cosine similarity 对大小不敏感，对方向敏感。使用卷积方便判断优化情况。
  - p=Duuv, C is a tensor of rank 4.

Materials and Design

- **Source**: British National Corpus (BNC).
- **Tool**: RASP (a syntactic analyzer), WordNet with Lesk similarity.
- **Extracted phrase type**: **adjective–noun**, **noun–noun**, and **verb–object** combinations.
- **Word representation**: (2 kinds in this paper.)
  - **Co-occurrence based semantic space**. Each vector component related to a particular context word, whose value is based on its co-occurrence with the target.
  - **Probabilistic topic models**. LDA topic model (omitted.)

Composition functions

Composition functions considered in our experiments

Model	Function
Additive	$p_i = u_i + v_i$
Kintsch	$p_i = u_i + v_i + n_i$
Multiplicative	$p_i = u_i \cdot v_i$
Tensor product	$p_{i,j} = u_i \cdot v_j$
Circular convolution	$p_i = \sum_j u_j \cdot v_{i-j}$
Weighted additive	$p_i = \alpha v_i + \beta u_i$
Dilation	$p_i = v_i \sum_j u_j u_j + (\lambda - 1) u_i \sum_j u_j v_j$
Head only	$p_i = v_i$
Target unit	$p_i = v_i(t_1 t_2)$

Evaluation & Results

- **Words**: a1, b1, a2, b2
- **Composition function**: *f*(x,y)
- **Phrases**: p1, p2
- p1 = *f*(a1, b1), p2=*f*(a2,b2)
- Calculate *sim*(p1, p2)

- **subject ratings** (7 points rating scale) vs. **model predictions**
- Use **correlation analysis** (**Spearman's ρ correlation coefficient,** 正相关+1 负相关-1) to examine the relationship between the **human ratings** and **vector-based similarity values**.

Results & Discussion

Co-occurrence based semantic space

Correlation coefficients of model predictions with subject similarity ratings (Spearman's ρ) using a simple semantic space

Model	Adjective–Noun	Noun–Noun	Verb–Object
Additive	.36	.39	.30
Kintsch	.32	.22	.29
Multiplicative	.46	.49	.37
Tensor product	.41	.36	.33
Convolution	.09	.05	.10
Weighted additive	.44	.41	.34
Dilation	.44	.41	.38
Target unit	.43	.34	.29
Head only	.43	.17	.24
Humans	.52	.49	.55

* **Spearman's ρ correlation coefficient** 大约是正相关+1、负相关-1。

- 所有模型预测结果均与人类判断正相关，其中循环卷积效果最差。
- Additive 类模型中，Kintsch 并没有对Additive 产生改进。
- Multi 类模型比 Additive 类型效果好。
- 在所有的 Multi 类模型中，简单的Multiplicative 比其他的Tensor prod、Conv 都要好。
- 总之，Multi 类模型、Additive 类模型、Dilation 表现优异。但都落后于人类。

Results & Discussion

Probabilistic topic models - LDA

Correlation coefficients of model predictions with subject similarity ratings (Spearman's ρ) using the LDA topic model

Model	Adjective–Noun	Noun–Noun	Verb–Object
Additive	.37	.45	.40
Kintsch	.30	.28	.33
Multiplicative	.25	.45	.34
Tensor product	.39	.43	.33
Convolution	.15	.17	.12
Weighted additive	.38	.46	.40
Dilation	.38	.45	.41
Head only	.35	.27	.17
Humans	.52	.49	.55

* **Spearman's ρ correlation coefficient** 大约是正相关+1、负相关-1。

- 所有模型预测结果均与人类判断正相关，其中循环卷积效果最差。
- Additive 类模型中，Kintsch 并没有对Additive 产生改进。
- Additive、Weighted Add、Dilation 表现类似。
- Tensor 在 AN 上显著胜过简单的 Multi，在 NN、VO 上持平。
- 和前面语义空间模型比，LDA 表现略差，归咎于其稀疏性。

Thank You!