Distributional Semantics: Applications, Structured Models

Pawan Goyal

CSE, IIT Kharagpur

Week 7, Lecture 3

Application to Query Expansion: Addressing Term Mismatch

Term Mismatch Problem in Information Retrieval

- Stems from the word independence assumption during document indexing.
- User query: *insurance cover which pays for long term care.*
- A relevant document may contain terms different from the actual user query.
- Some relevant words concerning this query: $\{medicare, premiums, insurers\}$

Application to Query Expansion: Addressing Term Mismatch

Term Mismatch Problem in Information Retrieval

- Stems from the word independence assumption during document indexing.
- User query: *insurance cover which pays for long term care.*
- A relevant document may contain terms different from the actual user query.
- Some relevant words concerning this query: $\{medicare, premiums, insurers\}$

Using DSMs for Query Expansion

Given a user query, reformulate it using related terms to enhance the retrieval performance.

- The distributional vectors for the query terms are computed.
- Expanded query is obtained by a linear combination or a functional combination of these vectors.

Query Expansion using Unstructured DSMs

TREC Topic 104: *catastrophic health insurance*

Query Representation: surtax:1.0 hcfa:0.97 medicare:0.93 hmos:0.83 medicaid:0.8 hmo:0.78 beneficiaries:0.75 ambulatory:0.72 premiums:0.72 hospitalization:0.71 hhs:0.7 reimbursable:0.7 deductible:0.69

Query Expansion using Unstructured DSMs

TREC Topic 104: catastrophic health insurance

Query Representation: surtax:1.0 hcfa:0.97 medicare:0.93 hmos:0.83 medicaid:0.8 hmo:0.78 beneficiaries:0.75 ambulatory:0.72 premiums:0.72 hospitalization:0.71 hhs:0.7 reimbursable:0.7 deductible:0.69

- Broad expansion terms: **medicare, beneficiaries, premiums** ...
- Specific domain terms: **HCFA** (Health Care Financing Administration), **HMO** (Health Maintenance Organization), **HHS** (Health and Human Services)

Query Expansion using Unstructured DSMs

TREC Topic 104: *catastrophic health insurance*

Query Representation: surtax:1.0 hcfa:0.97 medicare:0.93 hmos:0.83 medicaid:0.8 hmo:0.78 beneficiaries:0.75 ambulatory:0.72 premiums:0.72 hospitalization:0.71 hhs:0.7 reimbursable:0.7 deductible:0.69

- Broad expansion terms: **medicare, beneficiaries, premiums** ...
- Specific domain terms: **HCFA** (Health Care Financing Administration), **HMO** (Health Maintenance Organization), **HHS** (Health and Human Services)

TREC Topic 355: *ocean remote sensing*

Query Representation: radiometer:1.0 landsat:0.97 ionosphere:0.94 cnes:0.84 altimeter:0.83 nasda:0.81 meterology:0.81 cartography:0.78 geostationary:0.78 doppler:0.78 oceanographic:0.76

Query Expansion using Unstructured DSMs

TREC Topic 104: catastrophic health insurance

Query Representation: surtax:1.0 hcfa:0.97 medicare:0.93 hmos:0.83 medicaid:0.8 hmo:0.78 beneficiaries:0.75 ambulatory:0.72 premiums:0.72 hospitalization:0.71 hhs:0.7 reimbursable:0.7 deductible:0.69

- Broad expansion terms: **medicare, beneficiaries, premiums** ...
- Specific domain terms: **HCFA** (Health Care Financing Administration), **HMO** (Health Maintenance Organization), **HHS** (Health and Human Services)

TREC Topic 355: ocean remote sensing

Query Representation: radiometer:1.0 landsat:0.97 ionosphere:0.94 cnes:0.84 altimeter:0.83 nasda:0.81 meterology:0.81 cartography:0.78 geostationary:0.78 doppler:0.78 oceanographic:0.76

- Broad expansion terms: **radiometer, landsat, ionosphere** ...
- Specific domain terms: **CNES** (Centre National dÉtudes Spatiales) and **NASDA** (National Space Development Agency of Japan)

Similarity Measures for Binary Vectors

Let X and Y denote the binary distributional vectors for words X and Y.

Similarity Measures

Dice coefficient : $\frac{2|X \cap Y|}{|X|+|Y|}$

Similarity Measures for Binary Vectors

Let X and Y denote the binary distributional vectors for words X and Y.

Similarity Measures

Dice coefficient : $\frac{2|X \cap Y|}{|X| + |Y|}$

Jaccard Coefficient : $\frac{|X \cap Y|}{|X \cup Y|}$

Similarity Measures for Binary Vectors

Let X and Y denote the binary distributional vectors for words X and Y.

Similarity Measures

Dice coefficient : $\frac{2|X \cap Y|}{|X|+|Y|}$

Jaccard Coefficient : $\frac{|X \cap Y|}{|X \cup Y|}$

Overlap Coefficient : $\frac{|X \cap Y|}{min(|X|,|Y|)}$

Similarity Measures for Binary Vectors

Let X and Y denote the binary distributional vectors for words X and Y.

Similarity Measures

$$\text{Dice coefficient : } \frac{2|X \cap Y|}{|X|+|Y|}$$

$$\text{Jaccard Coefficient : } \frac{|X \cap Y|}{|X \cup Y|}$$

$$\text{Overlap Coefficient : } \frac{|X \cap Y|}{min(|X|,|Y|)}$$

Jaccard coefficient penalizes small number of shared entries, while Overlap coefficient uses the concept of inclusion.

Similarity Measures for Vector Spaces

Let $\vec{X}$ and $\vec{Y}$ denote the distributional vectors for words X and Y.
$\vec{X} = [x_1, x_2, \ldots, x_n]$, $\vec{Y} = [y_1, y_2, \ldots, y_n]$

Similarity Measures for Vector Spaces

Let $\vec{X}$ and $\vec{Y}$ denote the distributional vectors for words X and Y.
$\vec{X} = [x_1, x_2, \ldots, x_n]$, $\vec{Y} = [y_1, y_2, \ldots, y_n]$

Similarity Measures

Cosine similarity : $cos(\vec{X}, \vec{Y}) = \frac{\bar{X} \cdot \vec{Y}}{|\vec{X}||\vec{Y}|}$

Similarity Measures for Vector Spaces

Let $\vec{X}$ and $\vec{Y}$ denote the distributional vectors for words X and Y.
$\vec{X} = [x_1, x_2, \ldots, x_n]$, $\vec{Y} = [y_1, y_2, \ldots, y_n]$

Similarity Measures

Cosine similarity : $cos(\vec{X}, \vec{Y}) = \frac{\vec{X} \cdot \vec{Y}}{|\vec{X}||\vec{Y}|}$

Euclidean distance : $|\vec{X} - \vec{Y}| = \sqrt{\Sigma_{i=1}^{n} (x_i - y_i)^2}$

Similarity Measure for Probability Distributions

Let p and q denote the probability distributions corresponding to two distributional vectors.

Similarity Measure for Probability Distributions

Let p and q denote the probability distributions corresponding to two distributional vectors.

Similarity Measures

$$\text{KL-divergence}: D(p||q) = \Sigma_i p_i log \frac{p_i}{q_i}$$

$$\text{Information Radius}: D(p||\frac{p+q}{2}) + D(q||\frac{p+q}{2})$$

$$L_1\text{-norm}: \Sigma_i |p_i - q_i|$$

Attributional Similarity vs. Relational Similarity

Attributional Similarity

The attributional similarity between two words a and b depends on the degree of correspondence between the properties of a and b.
Ex: dog and wolf

Relational Similarity

Two pairs(a, b) and (c, d) are relationally similar if they have many similar relations.
Ex: dog: bark and cat: meow

Relational Similarity: Pair-pattern matrix

Pair-pattern matrix

- Row vectors correspond to pairs of words, such as *mason: stone* and *carpenter: wood*
- Column vectors correspond to the patterns in which the pairs occur, e.g. *X cuts Y* and *X works with Y*
- Compute the similarity of rows to find similar pairs

Relational Similarity: Pair-pattern matrix

Pair-pattern matrix

- Row vectors correspond to pairs of words, such as *mason: stone* and *carpenter: wood*
- Column vectors correspond to the patterns in which the pairs occur, e.g. *X cuts Y* and *X works with Y*
- Compute the similarity of rows to find similar pairs

Extended Distributional Hypothesis; Lin and Pantel

Patterns that co-occur with similar pairs tend to have similar meanings.

Relational Similarity: Pair-pattern matrix

Pair-pattern matrix

- Row vectors correspond to pairs of words, such as *mason: stone* and *carpenter: wood*
- Column vectors correspond to the patterns in which the pairs occur, e.g. *X cuts Y* and *X works with Y*
- Compute the similarity of rows to find similar pairs

Extended Distributional Hypothesis; Lin and Pantel

Patterns that co-occur with similar pairs tend to have similar meanings.
This matrix can also be used to measure the semantic similarity of patterns.

Relational Similarity: Pair-pattern matrix

Pair-pattern matrix

- Row vectors correspond to pairs of words, such as *mason: stone* and *carpenter: wood*
- Column vectors correspond to the patterns in which the pairs occur, e.g. *X cuts Y* and *X works with Y*
- Compute the similarity of rows to find similar pairs

Extended Distributional Hypothesis; Lin and Pantel

Patterns that co-occur with similar pairs tend to have similar meanings.
This matrix can also be used to measure the semantic similarity of patterns.
Given a pattern such as "X solves Y", you can use this matrix to find similar patterns, such as "Y is solved by X", "Y is resolved in X", "X resolves Y".

Structured DSMs

Basic Issue

- Words may not be the basic context units anymore
- How to capture and represent syntactic information?
  X solves Y and *Y is solved by X*

Structured DSMs

Basic Issue

- Words may not be the basic context units anymore
- How to capture and represent syntactic information?
  X solves Y and *Y is solved by X*

An Ideal Formalism

- Should mirror semantic relationships as close as possible
- Incorporate word-based information and syntactic analysis
- Should be applicable to different languages

Structured DSMs

Basic Issue

- Words may not be the basic context units anymore
- How to capture and represent syntactic information?
  X solves Y and *Y is solved by X*

An Ideal Formalism

- Should mirror semantic relationships as close as possible
- Incorporate word-based information and syntactic analysis
- Should be applicable to different languages

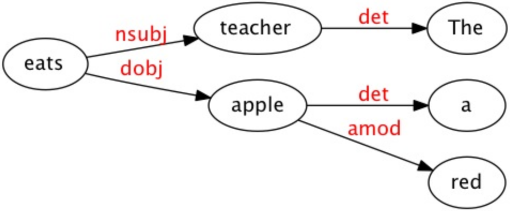
Use Dependency grammar framework

Structured DSMs

Structured DSMs

Using Dependency Structure: How does it help?

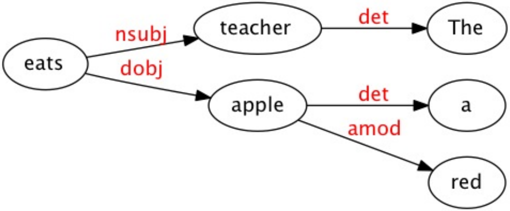
The teacher eats a red apple.



Structured DSMs

Using Dependency Structure: How does it help?

The teacher eats a red apple.



- 'eat' is not a legitimate context for 'red'.
- The 'object' relation connecting 'eat' and 'apple' is treated as a different type of co-occurrence from the 'modifier' relation linking 'red' and 'apple'.

Structured DSMs

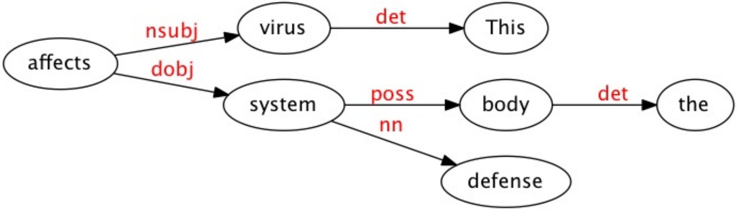
Structured DSMs: Words as 'legitimate' contexts

- Co-occurrence statistics are collected using parser-extracted relations.
- To qualify as context of a target item, a word must be linked to it by some (interesting) lexico-syntactic relation

Structured DSMs

Distributional models, as guided by dependency

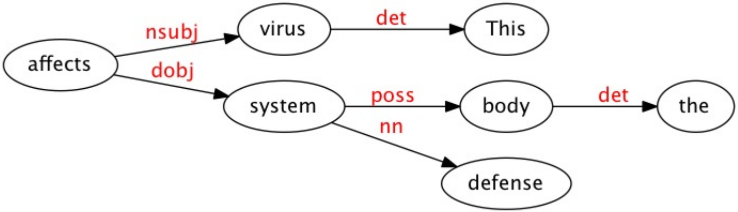
Ex: For the sentence '*This virus affects the body's defense system.*', the dependency parse is:



Structured DSMs

Distributional models, as guided by dependency

Ex: For the sentence '*This virus affects the body's defense system.*', the dependency parse is:



Word vectors

<system, dobj, affects> ...

Corpus-derived ternary data can also be mapped onto a 2-way matrix

2-way matrix

<system, dobj, affects>
<virus, nsubj, affects>

The dependency information can be dropped

- <system, dobj, affects> ⇒ <system, affects>
- <virus, nsubj, affects> ⇒ <virus, affects>

2-way matrix

<system, dobj, affects>
<virus, nsubj, affects>

The dependency information can be dropped

- <system, dobj, affects> ⇒ <system, affects>
- <virus, nsubj, affects> ⇒ <virus, affects>

Link and one word can be concatenated and treated as attributes

- *virus*={nsubj-affects:0.05,...},
- *system*={dobj-affects:0.03,...}

Structured DSMs for Selectional Preferences

Selectional Preferences for Verbs

Most verbs prefer arguments of a particular type. This regularity is known as selectional preference.

Structured DSMs for Selectional Preferences

Selectional Preferences for Verbs

Most verbs prefer arguments of a particular type. This regularity is known as selectional preference.

- From a parsed corpus, noun vectors are calculated as shown for ‘virus’ and ‘system’.

Structured DSMs for Selectional Preferences

Selectional Preferences for Verbs

Most verbs prefer arguments of a particular type. This regularity is known as selectional preference.

- From a parsed corpus, noun vectors are calculated as shown for ‘virus’ and ‘system’.

	obj-carry	obj-buy	obj-drive	obj-eat	obj-store	sub-fly	...
car	0.1	0.4	0.8	0.02	0.2	0.05	...
vegetable	0.3	0.5	0	0.6	0.3	0.05	...
biscuit	0.4	0.4	0	0.5	0.4	0.02	...
...	...	...	...	...	...	...	...

Selectional Preferences

- Suppose we want to compute the selectional preferences of the nouns as object of verb ‘eat’.

Structured DSMs for Selectional Preferences

Selectional Preferences

- Suppose we want to compute the selectional preferences of the nouns as object of verb ‘eat’.
- n nouns having highest weight in the dimension ‘obj-eat’ are selected, let {vegetable, biscuit,...} be the set of these n nouns.

Selectional Preferences

- Suppose we want to compute the selectional preferences of the nouns as object of verb 'eat'.
- n nouns having highest weight in the dimension 'obj-eat' are selected, let $\{$vegetable, biscuit,$\ldots\}$ be the set of these n nouns.
- The complete vectors of these n nouns are used to obtain an 'object prototype' of the verb.

Structured DSMs for Selectional Preferences

Selectional Preferences

- Suppose we want to compute the selectional preferences of the nouns as object of verb 'eat'.
- n nouns having highest weight in the dimension 'obj-eat' are selected, let $\{$vegetable, biscuit,$\ldots\}$ be the set of these n nouns.
- The complete vectors of these n nouns are used to obtain an 'object prototype' of the verb.
- 'object prototype' will indicate various attributes such as these nouns can be consumed, bought, carried, stored etc.

Structured DSMs for Selectional Preferences

Selectional Preferences

- Suppose we want to compute the selectional preferences of the nouns as object of verb ‘eat’.
- n nouns having highest weight in the dimension ‘obj-eat’ are selected, let $\{$vegetable, biscuit,$\ldots\}$ be the set of these n nouns.
- The complete vectors of these n nouns are used to obtain an ‘object prototype’ of the verb.
- ‘object prototype’ will indicate various attributes such as these nouns can be consumed, bought, carried, stored etc.
- Similarity of a noun to this ‘object prototype’ is used to denote the plausibility of that noun being an object of verb ‘eat’.