The PoP Approach to Vector Semantics

Laura Kallmeyer and Behrang QasemiZadeh

Semantics and Pragmatics Exchange Colloquium

June 15, 2016
Structure

• Introduction
  • The Distributional Hypothesis and Vector Space Models
• Problem Statement and Motivation
• A Solution: The PoP Method
• Reports from Experiments
• Conclusion
Harris’s Distributional Structure

• Does language have a distributional structure?
  • “The parts of a language do not occur arbitrarily relative to each other: each element occurs in certain positions relative to certain other elements.”
  • “All elements in a language can be grouped into classes whose relative occurrence can be stated exactly.”

• Meaning as a function of distribution.

Harris (1954) Distributional structure, WORD, 10:2-3, 146—162.
Harris’s Distributional Structure

• Does language have a distributional structure?
  • “The parts of a language do not occur arbitrarily relative to each other: each element occurs in certain positions relative to certain other elements.”
  • “All elements in a language can be grouped into classes whose relative occurrence can be stated exactly.”

• Meaning as a function of distribution.

• Language is not random.

Harris (1954) Distributional structure, WORD, 10:2-3, 146—162.
The Distributional Hypothesis

• Main idea: words with a similar meaning tend to occur in similar contexts.
The Distributional Hypothesis

• Main idea: words with a **similar meaning** tend to occur in **similar contexts**.

• Harris (1954) pointed out that “oculist and eye-doctor . . . occur in almost the same environment”.

• Most famous formulation of this idea goes back to Firth (1957): “You shall know a word by the company it keeps”.

• Simply Put, “the meaning of a word is related to the distribution of words around it (Jurafsky & Martin, 2015).”
The Distributional Hypothesis

• Main idea: words with a **similar meaning** tend to occur in **similar contexts**.

---

Example from Nida (1975); Lin (1998); Jurafsky & Martin (2015)

(1) a. A bottle of *tesgüino* is on the table.
   b. Everybody likes *tesgüino*.
   c. *Tesgüino* makes you drunk.
   d. We make *tesgüino* out of corn.
The Distributional Hypothesis (Example)

• Assume we want to investigate the meaning of two words “cherry” and “peach” using the distributional hypothesis.
The Distributional Hypothesis
(Example)

A base oil such as almond, grapeseed, peach or apricot kernel -- preferably cold-pressed before he told me that in Africa the peach tree is nothing more than an ordinary weed. A peach tree! A weed! So I dreamed I was held, the Chinese Isles of Orchids and their peach blossom spring, such ideas as Eldorado. An oyster catcher piped aboard a peach orange sunset all of three o'clock in the Imaging their amazement as the single peach on the tree grows to enormous proportion-streaky beach boy With a bum like a peach, boy The gentleman does prefer a blonde he looked stunning in a beautifully cut peach coloured suit and matching shoes. The them. The fruit is rather like a small peach, but the hull dies off towards the end. The base of these apples, such as strawberry or peach, too cloaking. Sparkling red or white apple-pip kernels, like those of plum, peach and apricot, contain tiny quantities of fragrant chiselled flowers, as we knew peach, plum, and cherry blossom; and birds e king-sized bed with its lace-trimmed peach silk sheets. Oh to be able to fall into them and their eyes met. Astorre, wiping peach juice from his beard, threw down his mended -RRB-, washed and halved 1 fresh peach, washed and sliced 8 -- 10 ratafia biscuits arrange the strawberries, grapes, and peach in the base of an approximately 900ml behind the packing sheds. The sun like a peach at seven, furring the railway lines with its silk dress by Rose Coutts-Smith, -LRB- peach Surprise, # 650 -RRB-, matching cathe made -- Liz chose silk dupion in a pale peach colour instead of the satin of the original outfits too -- long skirts with bustles in peach, and high-collared long striped jackets was the size of a railway station, all peach marble and glass and gilded metal. The chrysanthemums, white spider chrysanthemums, peach spray carnations and yellow lilies, as fat fromage frais followed by 1 sliced peach and 2 teaspoons raisins. LUNCH: 4 crisp a smooth skinned variety of the fuzzy peach, and are usually a deeper, red-orange new. An average 175g -LRB- 6 oz -RRB- peach contains 55 calories; a nectarine around
The Distributional Hypothesis
(Example)

erfully exuberant explosive mouthful of cherry fruit, herbs and spices –LRB- # 7.20 –RRB– with all my available teeth. She does a cherry. We go right past and down the road another 100m. 'Plena'; Fig. 2, Dwarf American cherry, Prunus pumila L.; drawing by R. Lancaster. Fruit trees –LRB- almond, peach, rose, cherry –RRB– caused by various biotypes of cherry Road, which was lined with flowering cherry trees, a great attraction for visitors. Sometimes we offer other fruit pies such as cherry, apricot, blueberry, etc. The ingredients I desire you will come over & thin your cherry trees w. you promised me to have done more bites at this financially rewarding cherry? If she did, it was an existence fraught with feet of clay with a stunted flowering cherry tree that shaded the sandbagged air-raid shelter wall of her home just in front of the cherry tree. She stared into the camera as if the tadpoles round her ankles, the cherry blossoms which she had thrown in his face. Definitely non-alcoholic -- it's called cherry brandy. 'A tankard was filled with red but we had taken a couple of bites at the cherry on our approach -- I could only guess how lysts. 'Florid words fell like unwanted cherry pips, now to lie and bleach on stone. And buzzed like hot island bees on sweet cherry trees. Swimming in the cooling of aqua from the sea, so for the first time the cherry red glow of liquid lava will be seen, and the yellow glow gives way to a bright cherry red, and black streaks and blobs of chilli and curds. When the lava has reached this cherry red temperature, it's extremely viscous, hazel, holly, round-leaved willow, cherry –LRB- gean –RRB–, aspen, dog rose, juniper. Mustn't lakked doon it since Hung oan to her cherry jist to marry a Prince Wi jumbo size ears to be worked on later this year, and two cherry trees. The trees, they hope, will be "5 on Alton 86675. Planting one of their cherry trees at Rotherfield Park are the 3rd Alt 86675 room. He uses six native woods: yew, cherry, ash, lime, sycamore and walnut as well an identical soundboards, lutes made of cherry and rosewood will sound quite different. Sounds expert, extrovert playing while the cherry produces a much softer, less showy sound. Flowers, as we knew peach, plum, and cherry blossom; and birds from flamingoes, pe...
The Distributional Hypothesis
(Example)

The word “peach” co-occurred 9 times with the word “tree”.

The word “peach” co-occurred 4 times with the word “fruit”.

The word “cherry” co-occurred 4 times with the word “tree”.

The word “cherry” co-occurred 6 times with the word “fruit”.

I love all such as cherry, apple, peach, plum, apricot kernel - preferably sulphate at before he told me that he didn’t think the tree was kicking more than an ordinary’s string more than an ordinary one. A peach tree 1 saw 1 desired. I was told

the Chinese fly of orchids and their peach blossom spring, such bloss as England say.

An easy method is planted a peach orange sunset all of three inches in the beginning then amendment as the single peach on the tree grows to enormous prunus

nectarine beach boy with a brand like a peach. boy. The gentleman does purchase a box in the fruit is a beautifully cut peach coloured cut and watching them. The

then. The fruit is rather like a small peach, but the husk does off towards the end are with peaches, such as strawberries or peaches, too, tasting. Sparkling red or white

apple-pie-kernels, like those of plum, peach and apricot: contain tiny quantities of nigella blelled flowers, as we know peach, plum, and cherry blossoms; and birds

a long connected with its juice-narrow peach oil sheets. Do to be able to fall into

and their eye, wet, azure, flaming peach juice from his beard, threw down his

red robe, washed and heated 1 fresh peach, washed and stood 6 - 10 minutes below

enameled the strawberries; grapes; and peach in this bowl of an approximately 59ml.

used the peeling knife. The sun like a peach at springs, during the rainbow with

stiff stems by Aki-san Smith, LBR - peach pinsure 1.150 - 3000 - matching coat

male - bix choice silk damask in a pale peach colour instead of the palms of the king

juvenile two - long strips with boson in peach; and high collared long-sleeved

a pair of a salmon column; all peach and glass and gemstone metal. The

vines, white graham chrysanthemum... peach spray (chrysanthemum and pelargonium: all)

fat from the face followed by 1 smooth peaches and 2 neophyte whites. LONGH - 4 crts

(a smooth almond variety of the fuzzy peach), and are usually a deeper, melon-orange

red. An average 15g. LBR: 6 oz. peaches contains 55 calories; a nectarine aour
The Distributional Hypothesis

(Example)

The word “peach” co-occurred 9 times with the word “tree”.
The word “peach” co-occurred 4 times with the word “fruit”.

The word “cherry” co-occurred 4 times with the word “tree”.
The word “cherry” co-occurred 6 times with the word “fruit”.

Let’s formalize this information in a vector space model.
The Distributional Hypothesis

(Example)

I base all of such as shrub, grape, peach, apricot kernel — preferably seedless at
the summit, for all that I tell you about the peach tree is nothing more than an ordinary
sting more than an ordinary nectar. A peach tree is a strange, 1-shaped, 1-shaped, 1-shaped,
and the Chinese idea of orchids and their peach blossom spring, such ideas as Edelweiss
say. An exotic crash my way over the peach orange sunset at all in this shrub in the
beaming brightishment as the single peach on the tree grew to enormous proper
fruit, beefy, peach and boy. The gentleman does prefer a blar
he boxed staring in a beautifully cut peach coloured cut and watching them.

The. The fruit is either a small peach, but the hub does off towards the end
instand drinks, such as strawberry, apricot, rose, dressing, sparkling red or white
apple-perfumery, like those of apple, peach and apricot... contain tiny quantities of
magnificent yellow flowers, as we know peach, plum, and cherry blossom; and birds
and long contact with its own-tinted peach silk sheet. Do I dare to lie to
and their eye, feel, attribute, savoring peach juice from his hand, then down his n
red, must, washed and blown 1 peach, washed and blown 0 — 12 weeks too
change the strawberries; grapes; and peach in the box of an approximately 50cm
took the packing stage. The sun-like a peach at once, turning the strawberry tree with
still-dressed by Asian Caro-Smith, UPR, peach Surprises, P 130.000, — watching cattle
male 1 losing silk in a pale peach colour instead of the palms of the virgin
outfits too — long skirts with buttons in peach; and high collared long striped jackets
was the size of a railway station, all peach marble and glass and gilded metal — Th
vines, white ladder, white cherries, peach spray, bitters and potted vines; all
first rowing fans followed by 1 second peach and 2 weapons native. UPR - $41
(a smooth sliced) variety of the buzz peach, and are usually a deeper, more-amber
red. An average 178g UPR 8 oz. ARB — peach contains 15 calories; a neartaneous
artfully exuberant expression of cherry fruit, roots and sprout. UPR; P 119.00
with all my available teeth, she uses a cherry. We go right past and down the road as
m.. Plane Fig. 1, Dwarf American cherry. Round puffs, growing by R. Caro-
not three UPR, alarmed, peach, may. Cherry ARB caused by the various botanists of
by a field which was laced with flowering cherry trees, a great attraction for natives
here we offer other fruit pies such as cherry, apple, plums, strawberries... The ingred
pot I desire you will come over two tiny cherry trees you promised me to have done
more loose at this financially rewarding cherry bill, I did, it was an excellent thought
about 10 feet of silk with a startling flowering cherry tree that shaded the sandbagged area
by the road just in front of the cherry tree. She donned into the camera as if
and the locals round her armpit — the cherry blossoms which she had thrown into the pe
definitely cost-economic. — it is called cherry beauty. — A twisted was filled red
we had taken a couple of times at the cherry on our approach — I could only guess his
height. — Parent wants tall beautiful cherry, pips, how to be and blue. At once,
white bushed like bolt felled bares or sweet cherry trees. Swimming in the cooling of aqua
from the sea, just for the first time the cherry red glow of nectar even be seen;
if the yellow glow gives way to a bright cherry red, black streaks and bands of red
result. Where the nectar has escaped into cherry red temperature, it is extremely vicious
under. Holly, round-leaved yellow, cherry UPR price $8.00, — apple now, ju
bloomed. Such a single tap on her cherry just to marry a Mexican little farmer can
be worked or later this year, and two cherry trees. The trees; they hope — will
be in color at all. Planting one of these cherry trees at Northfield Farm are the 3rd all
more; the trees as native winds now, cherry, eat, bite, pomegranate and mulberry as no
in chemical soundboards; lasts made of cherry and renowned wood sound quite different
as expert, excellent plying while the cherry produces a much softer, less showy cut
flowers, as we know peach, plum; and cherry blossom; and birds from flamingoes; pe
The Distributional Hypothesis

(Example)

I base all of this on input, grape, peach, or apricot kernel—preferably seedless—preferably seedless at first but told him that I had found the tree is lying more than an ordinary tree. A peach tree! I swear! I must have been told the Chinese are of orchids is a peach blossom spring, those things in Edgewood say. An easter, rather it showed a peach orange as at all three it stuck in the beginning then something else is the orange peach on the tree grows its enormous proper on a beach boy with a fruit—a peach, buy. The gentleman does prefer a bar he boxed standing in a beautifully cut peach colored cut and matching items. The thing! The fruit is either like a small peach, but the hub does off towards the end or instead drinks, such as strawberry or peach, too. Staying with sparkling red or white. All that, of course. The whole of those blossom. The entire of those blossom, peach and apricot. Though, tiny quantities of various colored flowers, as we know peach, plum, cherry blossom, and birds are long connected with its rococo turned peach seeds. It is to be able to fit into had his eye, tet, colloquially, saying peach juice from his beard, chewed down his red 10:00, washed and have 1 peach, washed and dried a 1/2 moons lens orange the strawberries, grapes; and peach in his box of an approximately. A fruit tied the packing sheet. The sun-like a peach at zero, turning the salad, then with still drier by Asia Coast-Smith, U.S., peach surprise. A 150:000, watching cattle make it. This choice is taken in a pale peach color instead of the scale of the pound to the limb—long strips with buds in peach, and high collared long-sleeved jackets was the size of a railway station, all peach material and glass and global metal. The violets, white jelly, cinnamon—peach, apricot, sour cherry, and yellow, etc. An all foliage fans followed by 1 small peach and 2 meadows, wild. The cherry, 9 smog (a smooth, similar). A metamorphic wholly of the fuzzy peach, and are usually a deeper, melon-color. An average 175g. 1 U.S. cat 488: peach contains 15.5 calories, a nectarine as
The Distributional Hypothesis (Example)

I base all of this on chords, grapes, peach or apricot kernels—preferably unpeeled at the bottom. I believe that the peach tree is higher than ordinary planting. A peach tree! A peach tree! I dreamed of it, I tell you. I dreamed of it! The Chinese love of fish and their peach blossom spring, such bliss as Eden seals!

An oyster eater poked aloud at a peach orange sunset at all three criticisms in the bringing them arrangement as the single peach on the tree grows its enormous prunus nectarine beach boyle with a bump like a peach. Boy! The gentleman does a bit of fruit tree blossoming in a beautiful peach coloured suit and matching shoes. The smell! The fruit is rather like a small peach, but the hull does off towards the end seashore drunks, such as strawberry or peach, blueberry, sparkling red or white smile-prune-kernels, like those of prune, peach and apricot. Content any quantity of native yellow flowers, as we know peach, plum, and cherry blossom; and birds are long connected with its ace-tinted peach pink leaves. To be able to fall into, and their eyes, feel, visualize, seeing peach juice from his beard, chew down his red lips, wash and believe. A peach in the first bed of an appropriately sized bowl, the peeling, the peeling. The sun like a peach at sunset, turning the grey clay with still deeper by Rosa Curtis-Smith, UPR peach Surprise, R-330-150, watching cattails—his case closed, it in a pale peach colour instead of the palms of the origin suitable for—long skirts without bustles in peach; and high collared long sleeved jackets was the size of a railway station, all peach marvelous and glass and grained metal. Th views, white paper (cigarette-paper): peach spray (cigarette-paper) and motion films; all—fare from street fans followed by 1 slim peach and 2 lemons insides. UNIFIED: + 4 tins (a smooth sided version of the fuzzy peach), and are usually a deeper, almost orange red. An average 175g or 6 oz ABB peach contains 55 calories: a nectarine about
The Distributional Hypothesis
(Example)

I base all of this on shorn ... peach or apricot kernel — preferably unskinned at before he told me that he arrived. the peach tree is nothing more than an ordinary ... string more than an ordinary one. A peach tree! I swear to do. I dream of ... she said. ... the Chinese love of orchids and their peach blossom spring — such bloss in Edelweiss say. An elderly lady placed a basket of orange uff as at all trying hard to bring the dim room back to the single peach on the tree grows its enormous protract n staring. beach boy with a head like a peach on. The gentleman does prefer a ... the tree Meyers. The tree became a basket of yellow and red. The tree grows its enormous projecting of peaches, as we know, peach, plum, and cherry blossoms; and birds e longer coiled with its narrow-tipped peach thicket sheets. To do to be able to fall into ... and their eyes, wet, wide-eyed, rising peach juice from his hands. Three dozen le ... red 6000, washed and blended 1 fresh peach, washed and stored 6 — 12 months less ... canteen, peaches, grapes, and peach in the box of an approximately 50cm2 ... the packing shed. The sun-tan of a peach at noon, turning the salmon color with ... the salary cutter, all peach mortar and glass and peeled metal. To video, white tablecloth stamped: peach syrup, mixture, and peach slices. The peaches were followed by 1ed peach and 2 pea ... 4 cups of smooth sliced carrots of the fuzzy peach, and are usually a deeper, ... orange 20. An average 17.5g U.S.-oz. 480B. peach contains 15 calories; a nectarine as well as ...
The Distributional Hypothesis

(Example)

I have all of such as shrimps, grapes...peach or apricot kernels—preferably wild apricot kernels before I told me that I think the peach tree is fruiting more than an ordinary year's string more than an ordinary year. A peach tree! I swear to I dreamed I was told...of the Chinese love of oranges and their peach blossom spring, such licks as Elijah says...An easter ushered this am with an orange sunset at all times when the blossoms then emasculated as the same peach on the tree gives the enormous pears...nearly beach boy with a burn like a peach...boy. The gentleman does prefer a bio...beak browsing in a beautifully cut peach colored suit and watching them. The thaw: The fruit is rather like a small peach...but the bush does off to the end of the season, such as strawberry or peach...too, peaches, sparkling red or white...apple pie kernels like those of apple, peach and apricot...contain tiny quantities of migrant chlorinated flowers, as we know...peach, plum, and cherry blossoms; and birds a long connected with its scarce trimmed peach orchards. Do we to be able to see into...and their eyes, even. I zone...saying peach juice from his hands...there again his red 0000...vowed and believed. I first peaches...vowed and sworn. I 0000...12 months less...orange the strawberries...pears; and peach in this box of approximately 5000...turtle the packing stick...The sun like a peach at eyes, turning the rail...their...swept by Asia Smith...LMR: peach Sunrise...LMR: 750 000...watching cattle make...as those who shiver in a pale peach color instead of the palms of the original...such--as long sticks with bolts in peach...and high collared long striped jackets...was the size of a railway...all peach mantle and glass and potted metal...The view...white paper (disembargers)...peach spray (compasses) and golden filler...all...fire from the fires followed by 1 smoked peaches and 2 apples in vases...LMR: 400...A smooth sense (smeared) of the fuzzy, peach...and are usually a deeper...maize orange box. An average 176g LMR: n 000--89--peach contains 55 calories...a nocturnal hour...artfully evocative expression of cherry fruit...berries and spices LMR: 7 29 89...with all my available teeth...she does a cherry...we go right past and down the main road...Page...Fig. 3, Dwarf American cherry. (Red...pincushion...driven by &...orange tail fines LMR: ahmed...peach. I now...cherry...LMR: caused by the various botanists of a...head, which was blessed with flowering cherry trees. They might for...i love we never other fruit past such as cherry...apricot...plum...etc...The ingrid...part I desire you will come over to this cherry tree. I was promised now...to have done same time...this financially-revealing cherry...if I did...it was an expedient brought...a host of folk with a freshly flowering cherry tree that stood the undamaged arms of the well of her home just in front of the cherry tree...She doused into the camera and...and the lollipops round her weak...the cherry blossoms which she had thrown into the form of the...not deficiently...cost-competitive...it is called cherry brandy...A lassow was filled with red...i saw this couple of times at the cherry on our approach...I could only guess his height...I saw one tall unvisited cherry; poppy...to be and behind me was...red but the lollipops...on sweet cherry trees. Swimming in the cooling of aqua...the sea...as for the first time the cherry radiated like liquid...will be seen...if the yellow glow gives way to a bright cherry red, and black streaks and blobs of oil...whatever when the lux has embraced the cherry red temperature, is an extremely viscous...half...hilly...Marked-leaved yellow...cherry...LMR: geen--89...acre...now...it looked down into the key to tap on her cherry...to marry. There is just one case which can be worked or save this year...and two cherry trees...The trees...they...hopes...it can still be felt...Raising one of those cherry trees at Rockefeller Park are the 300 dells...the trees...it is neither need...now...cherry...a long...trees...to the...mother and mother as it...intact soundboard...lutes made of cherry and recovered wood sound quite different...I'd expert...excellent playing while the cherry produces a much softer, less show real answers...as we know peach...plum...and cherry blossom...and birds from flummisters...pe
The Distributional Hypothesis
(Example)

Distance as a measure of similarity
The Distributional Hypothesis (Example)
The Distributional Hypothesis

(Example)
Summary, So Far!

• **Distributional Semantic Models** = the distributional hypothesis + vector space models.

  • Vector space models are used as a tool for formalizing the framework.
  • Vector space is a mathematical structure such as $\langle V, \mathbb{R}, +, \cdot \rangle$ that satisfy certain axioms:
    • A collection of vectors that can be added together and multiplied by numbers.
  • A notion of distance is used to compute similarities.
From Frequency to Meaning

- Hypothesis
- Collect Frequencies
- Vector Representation
- Compute Distances
- Results
From Frequency to Meaning: The Weighting Process

• The raw frequency counts are not the most informative measures of association between words:
  • Not all the co-occurrences are important to determine the meaning of a word;
  • Different co-occurrences have different impact on the meaning of words.
From Frequency to Meaning: The Weighting Process

• Association measures are used to convert raw frequencies into more informative numbers.

• Pointwise mutual information (PMI) is a well-known example: \( PMI(w, c) = \log_{2} \frac{P(w,c)}{P(w)P(c)} \).

• PMI gives us an estimate of how much more the word \( w \) and context word \( c \) co-occur than we would expect by chance.
From Frequency to Meaning

Hypothesis → Collect Frequencies → Vector Representation → Weighting → Compute Distances → Results
From Frequency to Meaning: Dimensionality Reduction

• Due to the *heavy-tailed* distribution of words, the number of context words escalates when new words are added to the model.

• Adding a new word requires adding additional dimensions to the vector space.
The dimensionality barrier

Dimensionality Increasing

100 words
The dimensionality barrier

Dimensionality Increasing

100 words

110 words
The dimensionality barrier

Dimensionality Increasing

- 100 words
- 110 words
- 120 words
The dimensionality barrier

Dimensionality Increasing

100 words

110 words

120 words

130 words

“THE CURSE OF DIMENSIONALITY”
The dimensionality barrier

• The Curse of Dimensionality:
  • Proportional impact of co-occurrences reduces when a model grows.
  • The meaning of two words is distinguishable only if they appear in completely different context.
  • In real world applications, models become computationally intractable due to their huge size.
From Frequency to Meaning

Hypothesis → Collect Frequencies → Vector Representation → Weighting → Dimension Reduction → Compute Distances → Results
Dimension Reduction
From sparse vectors to dense embeddings

• Truncation using singular value decomposition (SVD) is a classic method for reducing dimensionality.

• Idea:
  • Change the dimensions such that they are still orthogonal to each other.
  • Rank these new dimensions by the amount of variance in the data.
  • Instead of keeping all the dimensions resulting from this process, we only keep the top $k$ ones.
Dimension Reduction
From sparse vectors to dense embeddings

• Truncation using singular value decomposition (SVD) is a classic method for reducing dimensionality.
Dimension Reduction
From sparse vectors to dense embeddings

• Truncation using singular value decomposition (SVD) is a classic method for reducing dimensionality.

• Other popular methods for generating dense embeddings are skip-gram and continuous bag of words (CBOW).
Limitations of Matrix Factorization

• Computation of Singular Value Decomposition (SVD) is process-intensive, i.e. $O(mn^2)$:
  × Not suitable for big corpora!

• Every time a VSM is updated, SVD must be recomputed:
  × Not suitable for frequently updated text-data, e.g., applications involving text streams such as tweeter.
Limitations of Matrix Factorization

• Computation of Singular Value Decomposition (SVD) is process-intensive, i.e. $O(mn^2)$:
  × Not suitable for big corpora!

• Every time a VSM is updated, SVD must be recomputed:
  × Not suitable for frequently updated text-data, e.g., applications involving text streams such as tweeter.

**Solution:** Random Projection methods!
Random Projection Methods

• Using random projections, distributional models can be built directly at a reduced dimensionality without the need for computing SVD, etc.

• Dimensions of a model is decided independent of the number of context words and their distribution in the original vector space.
Random Projection Methods

- Random projections are often implemented as an incremental algorithm:
  - Map each target word to an empty “context vector”.
  - Map each context word to a randomly generated “index vector” (i.e., one index vector per context word).
  - Construct the VSM by accumulating index vectors to context vectors of co-occurring words.
Random Projection Methods (Example)

Assume we want to build a vector representation for the word “cherry”.

The taste of cherry is the best taste that I know.
Random Projection Methods
(Example)

Assume we want to build a vector representation for the word “cherry”.

The taste of cherry is the best taste that I know.

\[ [0, 0, 0, 0, 0, 0, 0, 0] \]
Random Projection Methods (Example)

Assume we want to built a vector representation for the word “cherry”.

The taste of cherry is the best taste that I know.

\[
\text{Vector(cherry)} = [2, 2, 1, 1, 1, 1, 1, 1, 1]
\]
Random Projection Methods (Example)

Assume we want to build a vector representation for the word “cherry”.

The taste of **cherry** is the **very** best taste that I know.

$$\text{Vector(cherry)} = [2, 2, 1, 1, 1, 1, 1, 1, 1, 1, 1]$$
Random Projection Methods (Example)

Assume we want to build a vector representation for the word “cherry”.

The taste of cherry is the best taste that I know.

\[
\text{Context Vector}(\text{cherry}) = [0, 0, 0, 0]
\]
Random Projection Methods (Example)

Assume we want to build a vector representation for the word “cherry”.

The taste of **cherry** is the best taste that I know.

<table>
<thead>
<tr>
<th>Context Vector(cherry)</th>
<th>[0, 0, 0, 0, 0]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Index Vector(the)</td>
<td>[-1, 0, 0, 1]</td>
</tr>
<tr>
<td>Index Vector(taste)</td>
<td>[0, 1, 0, -1]</td>
</tr>
<tr>
<td>Index Vector(of)</td>
<td>[-1, 1, 0, 0]</td>
</tr>
<tr>
<td>Index Vector(is)</td>
<td>[0, 0, 1, -1]</td>
</tr>
<tr>
<td>Index Vector(best)</td>
<td>[1, 0, -1, 0]</td>
</tr>
<tr>
<td>Index Vector(that)</td>
<td>[0, -1, 1, 0]</td>
</tr>
<tr>
<td>Index Vector(I)</td>
<td>[0, 1, -1, 0]</td>
</tr>
<tr>
<td>Index Vector(know)</td>
<td>[-1, 0, 1, 0]</td>
</tr>
</tbody>
</table>
Random Projection Methods
(Example)

Assume we want to build a vector representation for the word “cherry”.

The taste of cherry is the best taste that I know.

<table>
<thead>
<tr>
<th>Index Vector</th>
<th>Representation</th>
</tr>
</thead>
<tbody>
<tr>
<td>(the)</td>
<td>[-1, 0, 0, 1]</td>
</tr>
<tr>
<td>(taste)</td>
<td>[0, 1, 0, -1]</td>
</tr>
<tr>
<td>(of)</td>
<td>[-1, 1, 0, 0]</td>
</tr>
<tr>
<td>(is)</td>
<td>[0, 0, 1, -1]</td>
</tr>
<tr>
<td>(the)</td>
<td>[-1, 0, 0, 1]</td>
</tr>
<tr>
<td>(best)</td>
<td>[1, 0, -1, 0]</td>
</tr>
<tr>
<td>(taste)</td>
<td>[0, 1, 0, -1]</td>
</tr>
<tr>
<td>(that)</td>
<td>[0, -1, 1, 0]</td>
</tr>
<tr>
<td>(I)</td>
<td>[0, 1, -1, 0]</td>
</tr>
<tr>
<td>(know)</td>
<td>[-1, 0, 1, 0]</td>
</tr>
</tbody>
</table>
Random Projection Methods (Example)

Assume we want to build a vector representation for the word “cherry”.

The taste of cherry is the best taste that I know.

Index Vector (the) = [-1, 0, 0, 1]
Index Vector (taste) = [ 0, 1, 0, -1]
Index Vector (of) = [-1, 1, 0, 0]
Index Vector (is) = [ 0, 0, 1, -1]
Index Vector (the) = [-1, 0, 0, 1]
Index Vector (best) = [ 1, 0, -1, 0]
Index Vector (taste) = [ 0, 1, 0, -1]
Index Vector (that) = [ 0, -1, 1, 0]
Index Vector (I) = [ 0, 1, -1, 0]
Index Vector (know) = [-1, 0, 1, 0]

\[
\sum \text{Context Vector}(\text{cherry}) = [-3, 3, 1, -1]
\]
Random Projection Methods
(Example)

Assume we want to build a vector representation for the word “cherry”.

The taste of cherry is the best taste that I know.

Cherry = [ 2, 2, 1, 1, 1, 1, 1, 1, 1 ]
Cherry = [-3, 3, 1, -1]
Random Projection Methods (Example)

Assume we want to build a vector representation for the word “cherry”.

The taste of cherry is the very best taste that I know.

\[
\begin{align*}
\text{Index Vector}(\text{the}) &= [-1, 0, 0, 1] \\
\text{Index Vector}(\text{taste}) &= [0, 1, 0, -1] \\
\text{Index Vector}(\text{of}) &= [-1, 1, 0, 0] \\
\text{Index Vector}(\text{is}) &= [0, 0, 1, -1] \\
\text{Index Vector}(\text{the}) &= [-1, 0, 0, 1] \\
\text{Index Vector}(\text{best}) &= [1, 0, -1, 0] \\
\text{Index Vector}(\text{taste}) &= [0, 1, 0, -1] \\
\text{Index Vector}(\text{that}) &= [0, -1, 1, 0] \\
\text{Index Vector}(\text{I}) &= [0, 1, -1, 0] \\
\text{Index Vector}(\text{know}) &= [-1, 0, 1, 0] \\
\text{Index Vector}(\text{very}) &= [0, 0, -1, 1] \\
\end{align*}
\]

\[
\sum \text{Context Vector(cherry)} = [-3, 3, 0, 0]
\]
Random Projection Methods
(Example)

Assume we want to build a vector representation for the word “cherry”.

The taste of **cherry** is the **very** best taste that I know.

Index Vector *(the)* = [-1, 0, 0, 1]
Index Vector *(taste)* = [ 0, 1, 0,-1]
Index Vector *(of)* = [-1, 1, 0, 0]
Index Vector *(is)* = [ 0, 0, 1,-1]
Index Vector *(the)* = [-1, 0, 0, 1]
Index Vector *(best)* = [ 1, 0,-1, 0]
Index Vector *(taste)* = [ 0, 1, 0,-1]
Index Vector *(that)* = [ 0,-1, 1, 0]
Index Vector *(I)* = [ 0, 1,-1, 0]
Index Vector *(know)* = [-1, 0, 1, 0]
Index Vector *(very)* = [ 0, 0, -1,1]

Summation of context vectors:

\[
\sum \text{Context Vector(cherry)} = \begin{bmatrix} -3 \\ 3 \\ 0 \\ 0 \end{bmatrix}
\]
Random Projection Methods

• Random projection methods are not that random:
  • Randomly generated index vectors must obey certain distribution.
  • Depending on the distribution of index vectors, a suitable similarity (distance) measure must be chosen.
Random Projection Methods: Random Indexing (RI)

• In random indexing, index vectors have asymptotic standard Normal distribution:

\[ r_{ij} = \begin{cases} 
+1 & \text{With probability } \frac{s}{2} \\
0 & \text{With probability } 1 - s \\
-1 & \text{With probability } \frac{s}{2} 
\end{cases} \]

• For computing similarities in L2-normed spaces: cosine, Euclidean distance.
The shortcoming of RI

• Index vectors in RI (and similar techniques) have standard distribution:
  • The sum of coordinates in index vectors is always zero.
  • Hence, the sum of coordinates in context vectors is always zero too!

Context Vector(cherry) = \([-3, 3, 1, -1]\)
-3 + 3 + 1 + -1 = 0
The shortcoming of RI

• Index vectors in RI (and similar techniques) have standard distribution:
  • The sum of coordinates in index vectors is always zero.
  • Hence, the sum of coordinates in context vectors is always zero too!
• Many weighting techniques such as PMI cannot be used after the construction of models:
  • Simply, the problem of division by zero.
• As a result, these methods are not as good as techniques such as CBOW (the neural nets, etc.).
Positive-only Projection (PoP)

• PoP addresses this problem by introducing a new distribution for index vectors:

\[ r_{ij} = \begin{cases} \frac{1}{\sqrt{U}} & \text{With probability } s \\ 0 & \text{With probability } 1 - s \end{cases} \]

\( U \) is independent and identically distributed random variable in (0,1]

• Similarities between vectors is computed using Kendall’s correlation or Pearson’s Correlation.
The taste of cherry is the best taste that I know.

Index Vector (the) = [ 0, 0, 0, 1]
Index Vector (taste) = [ 0, 2, 0, 0]
Index Vector (of) = [ 0, 1, 0, 0]
Index Vector (is) = [ 0, 0, 3, 0]
Index Vector (the) = [ 0, 0, 0, 1]
Index Vector (best) = [ 1, 0, 0, 0]
Index Vector (taste) = [ 0, 2, 0, 0]
Index Vector (that) = [ 0, 2, 0, 0]
Index Vector (I) = [ 0, 0, 9, 0]
Index Vector (know) = [ 0, 0, 8, 0]

\[ \sum \text{Context Vector(cherry)} = [1, 7, 20, 2] \]
Advantages of PoP

• In PoP, the sum of coordinates in index vectors and thus context vectors is always greater than zero.

• Therefore, PoP context vectors can be weighted using association measures such as PMI after their construction:
  • Highly discriminative models can be built using small amount of computations.
Advantages of PoP

• It is much faster than matrix factorization techniques such as SVD truncation.

• Compared to deep learning methods:
  • PoP does not require any training.
  • PoP is much simpler to use:
    • In order to use deep learning methods, a number of parameters must be set prior to their training. However, PoP is almost a parameter-free method.
Evaluation

• To evaluate distributional methods, their performance is often assessed in tasks that requires modelling a notion of semantic similarity:

• **The TOEFL test** consists of a set of 80 questions, each consisting of a target word and 4 word choices:
  • Which of the following words has the closet meaning to the word “levied”:
    a) imposed, b) believed, c) requested, d) correlated.
Evaluation

• To evaluate distributional methods, their performance is often assessed in tasks that requires modelling a notion of semantic similarity:
  • The MEN relatedness test consists of 3000 pairs of words. Using a scale of 0 to 50, each pair is rated as semantically related by human:
    • sun-n sunlight-n 50.000000
    • automobile-n car-n 50.000000
    • river-n water-n 49.000000
    • ...
    • cafe-n frog-n 4.000000
    • bakery-n zebra-n 0.000000
Evaluation
PoP vs RI vs Standard Model vs CBOW (Word2Vec) in the TOEFL Test
Evaluation
PoP vs RI vs Standard Model vs CBOW (Word2Vec) in the MEN relatedness Test
Evaluation

- Increasing the dimensionality of models (up to a certain number) enhances their performance.
Current Application of Vector Space Models and PoP in B08
Word Sense Induction

• Word sense induction is modelled as a clustering tasks in distributional semantic models:
  • Group different occurrences of a lexical form (e.g., see) into a number of clusters.
  • Each cluster must represent one of the senses (meanings) of the lexical form.
  • The goal is to find sentences in which a word has the same meaning.
Example Sentences

Rev looked up and **saw** her.
Although I had been inside it I had not yet **seen** it functioning.
You **see** what I mean??
I mean I 've got to -- to **see** to the kitchen.
Chris was sound asleep , and I did n't **see** any sense in waking him.
He **saw** them ambushed , strewn in the postures of the broken and the dying.
Glad to **see** you go.

And it does n't **make** any difference.
He **made** a circle with his thumb and fingers.
Rifle leveled on the man , he **made** a rush.
Marlin has **made** two contributions to the harvest of new offerings.
There was not enough room to **make** the usual vertical bomb run.
By the time I 'd **made** it he was gone.
He 's mad at a world he did not **make**.
Example:
The output of clustering for two verbs see and make in the SemCor corpus.
Example:
The output of clustering for two verbs see and make in the SemCor corpus.

Each red box represents a cluster, i.e., a sense or a meaning, which are connected to each other in a hierarchy (blue lines).
What is our current research?

• We do not know which statistical information (among others) are the representative of words meanings in a sentence:
  • A systematic investigation of word co-occurrences, using information other than simple occurrences (e.g., using syntactic relationships, second-order occurrences, etc.).

• Addressing data sparseness: how to build a vector representation of a word only from a sentence!?  
  • Using methods such as PoP to draw meaningful conclusions from limited observations.
If interested to know more about PoP: