Word Space Models and Cognitive Research
State of the Art and Future Perspectives

Alessandro Lenci\textsuperscript{1} Marco Baroni\textsuperscript{2}

\textsuperscript{1}Department of Linguistics
University of Pisa

\textsuperscript{2}Center for Mind/Brain Sciences
University of Trento

Contextual Information in Word Space Models (CoSMo)
Roskilde, August 21 2007
Word Space Models (WSMs) are claimed to be plausible models of human knowledge organization and learning:

“The dimensionality–optimizing method offers a promising solution to the ancient puzzle of human knowledge induction. It still remains to determine how wide its scope is among human learning and cognition phenomena. [...] We would suggest that applications to problems in conditioning, association, pattern and object recognition, contextual disambiguation, metaphor, concepts and categorization, reminding, casebased reasoning, probability and similarity judgment, and complex stimulus generalization are among the set where this kind of induction might provide new solutions” (Landauer and Dumais 1997: 235)
WSMs and cognitive research

- Measures of semantic similarity based on WSMs have been demonstrated to predict behavioral performance in various tasks
  - synonymy identification (Landauer and Dumais 1997)
  - text coherence (Landauer and Dumais 1997)
  - categorization (Burgess and Lund 1997)
  - semantic priming in lexical decision tasks (Lowe 2000, McDonald and Brew 2002, Vigliocco et al. 2004)
  - word substitution errors (Vigliocco et al. 2004)
  - child vocabulary acquisition (Li et al. 2004)
  - etc.
WSMs and cognitive research

Six challenges for WSMs to be used as models of meaning and semantic memory

Incrementality

Multi-faceted semantic similarity

Interpretability and structure

Grounding

Context-sensitivity

Compositionality
Word learning occurs in an item-based way

- Children explore the word input space in a peacemeal fashion
- Semantic representations plausibly emerge out of local word co-occurrence distributions rather than from global statistical process that require the access to the whole input space
Incrementality as a cognitive constraint
Simulating word learning with WSMs

▶ *Time is important!*
  ▶ The way in which words affect each other evolves in time
  ▶ The information brought by seeing a word $c$ co-occurring with $w$ at time $t_i$ is different from the information brought by observing the same co-distribution at time $t_i + k$
    ▶ In this time interval, the semantics of $c$ and $w$ may be changed because of the other words they have independently encountered
How incremental are WSMs?

Latent Semantic Analysis (LSA) - Landauer and Dumais (1997)

- The family of models based on singular value decomposition (SVD) and similar dimensionality reduction techniques are very effective in building highly accurate semantic spaces (cf. Rapp 2006, et al.)
- SVD is claimed to be able to capture “second order” similarity relations (cf. Manning and Schütze 1999)
  - *first order*: two words are similar because they occur in the same context
  - *second order*: two words are similar because the contexts in which they occur are similar
How incremental are WSMs?
Latent Semantic Analysis (LSA) - Landauer and Dumais (1997)

- The family of models based on singular value decomposition (SVD) and similar dimensionality reduction techniques are very effective in building highly accurate semantic spaces (cf. Rapp 2006, et al.)
- SVD is claimed to be able to capture “second order” similarity relations (cf. Manning and Schütze 1999)
  - First order: two words are similar because they occur in the same context
  - Second order: two words are similar because the contexts in which they occur are similar
How incremental are WSMs?

Latent Semantic Analysis (LSA) - Landauer and Dumais (1997)

- Semantic accuracy is achieved by LSA at the expense of incrementality

- SVD is a global procedure that requires a full co-occurrence matrix based on statistics extracted from the whole input corpus

- However, Gorrell (2006) proposes an incremental version of SVD, based on hebbian learning
How incremental are WSMs?

Latent Semantic Analysis (LSA) - Landauer and Dumais (1997)

- Semantic accuracy is achieved by LSA at the expense of incrementality
- SVD is a global procedure that requires a full co-occurrence matrix based on statistics extracted from the whole input corpus
- However, Gorrell (2006) proposes an incremental version of SVD, based on hebbian learning
How incremental are WSMs?
Random Indexing (RI) - Karlgren and Sahlgren (2001)

- RI is incremental to the extent that vector representations are well-formed at each stage
- “Second order” effects may be missed
  - Since different random signatures are assigned to the words *cat*, *dog* and *train*, the model does not capture the fact that the first two words, but not the third, should count as similar contexts
  - RI is generally outperformed by LSA (Gorman and Curran 2006)
- “Weak” incrementality
  - The way in which a word affects another does not evolve in time
How incremental are WSMs?
Random Indexing (RI) - Karlgren and Sahlgren (2001)

- RI is incremental to the extent that vector representations are well-formed at each stage
- “Second order” effects may be missed
  - Since different random signatures are assigned to the words *cat, dog* and *train*, the model does not capture the fact that the first two words, but not the third, should count as similar contexts
  - RI is generally outperformed by LSA (Gorman and Curran 2006)
- “Weak” incrementality
  - The way in which a word affects another does not evolve in time
How incremental are WSMs?
Random Indexing (RI) - Karlgren and Sahlgren (2001)

- RI is incremental to the extent that vector representations are well-formed at each stage
- “Second order” effects may be missed
  - Since different random signatures are assigned to the words *cat, dog* and *train*, the model does not capture the fact that the first two words, but not the third, should count as similar contexts
  - RI is generally outperformed by LSA (Gorman and Curran 2006)
- “Weak” incrementality
  - The way in which a word affects another does not evolve in time
Incremental Semantic Analysis (ISA)
Baroni, Lenci and Onnis (2007)

- ISA is a fully incremental extension of RI
- A target word is affected not only by its context words, but also by the semantic information encoded by the distributional histories of these context words
  - ISA can capture SVD-like “second order” effects
    - *cat* and *dog* might work like similar contexts because they are likely to have similar histories.
- The information carried by a context word changes as our knowledge about the word increases
- ISA significantly outperforms SVD and RI in a word learning task, when incrementally trained on a small corpus (ca. 400K tokens) of child-directed speech
ISA is a fully incremental extension of RI

A target word is affected not only by its context words, but also by the semantic information encoded by the distributional histories of these context words

- ISA can capture SVD-like “second order” effects
  - *cat* and *dog* might work like similar contexts because they are likely to have similar histories.

The information carried by a context word changes as our knowledge about the word increases

ISA significantly outperforms SVD and RI in a word learning task, when incrementally trained on a small corpus (ca. 400K tokens) of child-directed speech
Incremental Semantic Analysis (ISA)
Baroni, Lenci and Onnis (2007)

- ISA is a fully incremental extension of RI
- A target word is affected not only by its context words, but also by the semantic information encoded by the distributional histories of these context words
  - ISA can capture SVD-like “second order” effects
    - *cat* and *dog* might work like similar contexts because they are likely to have similar histories.
- The information carried by a context word changes as our knowledge about the word increases
- ISA significantly outperforms SVD and RI in a word learning task, when incrementally trained on a small corpus (ca. 400K tokens) of child-directed speech
Incremental Semantic Analysis (ISA)
Baroni, Lenci and Onnis (2007)

- ISA is a fully incremental extension of RI
- A target word is affected not only by its context words, but also by the semantic information encoded by the distributional histories of these context words
  - ISA can capture SVD-like “second order” effects
    - *cat* and *dog* might work like similar contexts because they are likely to have similar histories.
- The information carried by a context word changes as our knowledge about the word increases
- ISA significantly outperforms SVD and RI in a word learning task, when incrementally trained on a small corpus (ca. 400K tokens) of child-directed speech
WSM and cognitive research

Six challenges for WSM to be used as models of meaning and semantic memory

Incrementality

Multi-faceted semantic similarity

Interpretability and structure

Grounding

Context-sensitivity

Compositionality
Multi-faceted semantic similarity

“Semantic similarity” is multi-faceted notion but a single WSM provides only one way to rank a set of words

Some nearest neighbours of *motorcycle* (output of WSM trained on BNC)

- motor → component
- car → co-hyponym
- diesel → component?
- to race → proper function
- van → co-hyponym
- bmw → hyponym
- to park → proper function
- vehicle → hypernym
- engine → component
- to steal → frame?
“Semantic similarity” is multi-faceted notion but a single WSM provides only one way to rank a set of words.

Some nearest neighbours of *motorcycle* (output of WSM trained on BNC):

- motor $\rightarrow$ component
- car $\rightarrow$ co-hyponym
- diesel $\rightarrow$ component?
- to race $\rightarrow$ proper function
- van $\rightarrow$ co-hyponym
- bmw $\rightarrow$ hyponym
- to park $\rightarrow$ proper function
- vehicle $\rightarrow$ hypernym
- engine $\rightarrow$ component
- to steal $\rightarrow$ frame?
Multi-faceted semantic similarity

- Different ways in which other words can be similar to a target word/concept:
  - Taxonomic relations (motorcycle and car)
  - Properties and parts of concept (motorcycle and engine)
  - Proper functions (motorcycle and to race)
  - Frame relations (motorcycle and to steal)
- Properties, parts, proper functions constitute representation of word/concept
- Ontological relations are product of overlapping representations in terms of properties etc.
  - A motorcycle is a motorcycle because it has an engine, two wheels, it is used for racing...
  - A car is similar to a motorcycle because they share a number of crucial properties and functions (engine and wheels, driving)
- This is not captured in WSM representation
Multi-faceted semantic similarity

- Different ways in which other words can be similar to a target word/concept:
  - Taxonomic relations (*motorcycle* and *car*)
  - Properties and parts of concept (*motorcycle* and *engine*)
  - Proper functions (*motorcycle* and *to race*)
  - Frame relations (*motorcycle* and *to steal*)
- Properties, parts, proper functions constitute representation of word/concept
- Ontological relations are product of overlapping representations in terms of properties etc.
  - A motorcycle is a motorcycle because it has an engine, two wheels, it is used for racing...
  - A car is similar to a motorcycle because they share a number of crucial properties and functions (engine and wheels, driving)
- This is not captured in WSM representation
Multi-faceted semantic similarity

- Different ways in which other words can be similar to a target word/concept:
  - Taxonomic relations (*motorcycle* and *car*)
  - Properties and parts of concept (*motorcycle* and *engine*)
  - Proper functions (*motorcycle* and *to race*)
  - Frame relations (*motorcycle* and *to steal*)

- Properties, parts, proper functions constitute *representation* of word/concept

- Ontological relations are product of overlapping representations in terms of properties etc.
  - A motorcycle is a motorcycle because it has an engine, two wheels, it is used for racing...
  - A car is similar to a motorcycle because they share a number of crucial properties and functions (engine and wheels, driving)

- This is not captured in WSM representation
Multi-faceted semantic similarity

- Different ways in which other words can be similar to a target word/concept:
  - Taxonomic relations (*motorcycle* and *car*)
  - Properties and parts of concept (*motorcycle* and *engine*)
  - Proper functions (*motorcycle* and *to race*)
  - Frame relations (*motorcycle* and *to steal*)

- Properties, parts, proper functions constitute representation of word/concept

- Ontological relations are product of overlapping representations in terms of properties etc.
  - A motorcycle is a motorcycle because it has an engine, two wheels, it is used for racing...
  - A car is similar to a motorcycle because they share a number of crucial properties and functions (engine and wheels, driving)

- This is not captured in WSM representation
WSM and cognitive research

Six challenges for WSM to be used as models of meaning and semantic memory

Incrementality

Multi-faceted semantic similarity

Interpretability and structure

Grounding

Context-sensitivity

Compositionality
Interpretability and structure

In WSM, word meaning is represented as a vector whose dimensions are:

- words, documents, etc.
- or, if dimensionality reduction technique is applied, “latent” factors

In either case, dimensions are hard/impossible to interpret

“As such, representations are purely abstract, denoting a word’s similarity to other words without revealing which aspects of meaning or relation are responsible in one case or another” (Vigliocco and Vinson 2007)
Interpretability and structure

- In WSM, word meaning is represented as a vector whose dimensions are:
  - words, documents, etc.
  - or, if dimensionality reduction technique is applied, “latent” factors
- In either case, dimensions are hard/impossible to interpret
  - “As such, representations are purely abstract, denoting a word’s similarity to other words without revealing which aspects of meaning or relation are responsible in one case or another” (Vigliocco and Vinson 2007)
Interpretability and structure

Converging evidence suggests that semantic representations are rich and complex structures of properties (= interpreted features)

- Rich lexical representations needed for semantic interpretation:
  - to finish a book (reading it) vs. an ice-cream (eating it) (Pustejovsky 1995)
  - a zebra pot is a pot with stripes

- Strong functional neuro-imaging evidence for property-based activation of sensory and motor systems (Martin 2007)

Some approaches try to provide more structure to WSMs:

- cf. also this workshop: Hagiwara et al., Peirsman et al., Van der Cruys
Interpretability and structure

- Converging evidence suggests that semantic representations are rich and complex structures of properties (interpreted features)
  - Rich lexical representations needed for semantic interpretation:
    - *to finish a book* (reading it) vs. *an ice-cream* (eating it) (Pustejovsky 1995)
    - a zebra pot is a pot with stripes
  - Strong functional neuro-imaging evidence for property-based activation of sensory and motor systems (Martin 2007)
- Some approaches try to provide more structure to WSMs:
  - cf. also this workshop: Hagiwara *et al.*, Peirsman *et al.*, Van der Cruys
Interpretability and structure

- Converging evidence suggests that semantic representations are rich and complex structures of properties (= interpreted features)
  - Rich lexical representations needed for semantic interpretation:
    - *to finish a book* (reading it) vs. *an ice-cream* (eating it) (Pustejovsky 1995)
    - a zebra pot is a pot with stripes
  - Strong functional neuro-imaging evidence for property-based activation of sensory and motor systems (Martin 2007)
- Some approaches try to provide more structure to WSMs:
  - cf. also this workshop: Hagiwara *et al.*, Peirsman *et al.*, Van der Cruys
Interpretability and structure

▶ Converging evidence suggests that semantic representations are rich and complex structures of properties (= interpreted features)
  ▶ Rich lexical representations needed for semantic interpretation:
    ▶ to finish a book (reading it) vs. an ice-cream (eating it) (Pustejovsky 1995)
    ▶ a zebra pot is a pot with stripes
  ▶ Strong functional neuro-imaging evidence for property-based activation of sensory and motor systems (Martin 2007)
▶ Some approaches try to provide more structure to WSMs:
  ▶ cf. also this workshop: Hagiwara et al., Peirsman et al., Van der Cruys
WSM and cognitive research

Six challenges for WSM to be used as models of meaning and semantic memory

Incrementality

Multi-faceted semantic similarity

Interpretability and structure

Grounding

Context-sensitivity

Compositionality
In current WSMs context is defined as “co-text”

The consequence is that semantic representations are not grounded in the extra-linguistic reality (cf. Glenberg and Robertson 2000, Vigliocco and Vinson 2007)

“By now many readers may wonder how the word similarities learned by LSA relate to meaning. Whereas it is probably impossible to say what word meaning is in a way that satisfies all students of the subject, it is clear that two of its most important aspects are usage and reference. Obviously, the similarity relations between words that are extracted by LSA are based on usage” (Landauer and Dumais 1997: 226-227)
Grounding

- In current WSMs context is defined as “co-text”
- The consequence is that semantic representations are not grounded in the extra-linguistic reality (cf. Glenberg and Robertson 2000, Vigliocco and Vinson 2007)
  - “By now many readers may wonder how the word similarities learned by LSA relate to meaning. Whereas it is probably impossible to say what word meaning is in a way that satisfies all students of the subject, it is clear that two of its most important aspects are usage and reference. Obviously, the similarity relations between words that are extracted by LSA are based on usage” (Landauer and Dumais 1997: 226-227)
Ungrounded semantic spaces

Pros:
- abstract concepts are most probably learnt through word usage
- many words (even concrete ones) are learnt (mostly or exclusively) via linguistic input

Cons:
- concepts (even abstract ones) can not be reduced to word co-occurrences
- strong behavioral and neuro-imaging evidence support the hypothesis that conceptual knowledge is deeply rooted in sensory-motor systems (cf. Barsalou 1999, 2005; Martin 2007)
Ungrounded semantic spaces

**Pros:**
- abstract concepts are most probably learnt through word usage
- many words (even concrete ones) are learnt (mostly or exclusively) via linguistic input

**Cons:**
- concepts (even abstract ones) can not be reduced to word co-occurrences
- strong behavioral and neuro-imaging evidence support the hypothesis that conceptual knowledge is deeply rooted in sensory-motor systems (cf. Barsalou 1999, 2005; Martin 2007)
Grounding WSMs

The challenge is to broaden the context of WSMs

▶ Which semantic properties of a word can be learnt from observing that word co-occurring with other words?
▶ Which semantic aspects of a word can be learnt from the perceptual features of its referent and/or of the situations in which it is normally found?

Mixed WSMs could eventually lead to interesting solutions to the “gavagai” paradox (Quine 1960)

▶ Cf. Howell et al. (2005) for a first proposal, but lots still remains to be done
The challenge is to broaden the context of WSMs

- Which semantic properties of a word can be learnt from observing that word co-occurring with other words?
- Which semantic aspects of a word can be learnt from the perceptual features of its referent and/or of the situations in which it is normally found?

Mixed WSMs could eventually lead to interesting solutions to the “gavagai” paradox (Quine 1960)

Cf. Howell et al. (2005) for a first proposal, but lots still remains to be done
WSM and cognitive research

Six challenges for WSM to be used as models of meaning and semantic memory

Incrementality

Multi-faceted semantic similarity

Interpretability and structure

Grounding

Context-sensitivity

Compositionality
WSMs in context

- Semantic representations in WSMs are **holistic** and **context-free**
  - Each word is represented by a single vector encoding its global co-occurrence history
  - Word vectors are invariant wrt different possible contexts of use of a word
- Words may have different meanings in different contexts
- Various attempts to carve senses out of word spaces to tackle disambiguation
  - cf. Schütze (1998), Pantel and Lin (2002), Van der Cruys (this workshop)
WSMs in context

- Semantic representations in WSMs are holistic and context-free
  - Each word is represented by a single vector encoding its global co-occurrence history
  - Word vectors are invariant wrt different possible contexts of use of a word
- Words may have different meanings in different contexts
- Various attempts to carve senses out of word spaces to tackle disambiguation
  - cf. Schütze (1998), Pantel and Lin (2002), Van der Cruys (this workshop)
Context-sensitive semantic spaces

- Context-sensitivity is a general feature of meaning (not limited to homonymy)
- Semantic similarity relations are context-dependent
  - *The school is boring* → school : book
  - *The school is out for a trip* → school : family
  - *The school was built in a year* → school : house
- Behavioral evidence supports the view that conceptual representations are inherently situated in contexts (cf. Barsalou 2005)
  - *dog* + zoological context → shares properties with *wolf*
  - *dog* + pet context → shares properties with *cat* or *parrot*
Context-sensitive semantic spaces

- Context-sensitivity is a general feature of meaning (not limited to homonymy)
- Semantic similarity relations are context-dependent
  - The school is boring $\rightarrow$ school : book
  - The school is out for a trip $\rightarrow$ school : family
  - The school was built in a year $\rightarrow$ school : house
- Behavioral evidence supports the view that conceptual representations are inherently situated in contexts (cf. Barsalou 2005)
  - dog + zoological context $\rightarrow$ shares properties with wolf
  - dog + pet context $\rightarrow$ shares properties with cat or parrot
Context-sensitive semantic spaces

- Context-sensitivity is a general feature of meaning (not limited to homonymy)
- Semantic similarity relations are context-dependent
  - *The school is boring* → school : book
  - *The school is out for a trip* → school : family
  - *The school was built in a year* → school : house
- Behavioral evidence supports the view that conceptual representations are inherently situated in contexts (cf. Barsalou 2005)
  - *dog* + zoological context → shares properties with *wolf*
  - *dog* + pet context → shares properties with *cat* or *parrot*
WSM and cognitive research

Six challenges for WSM to be used as models of meaning and semantic memory

Incrementality

Multi-faceted semantic similarity

Interpretability and structure

Grounding

Context-sensitivity

Compositionality
Beyond word meaning

- Few attempts to use WSMs to represent sentence meaning
- “Bag of words” approach (cf. Landauer and Dumais 1997)
  - Sentence (discourse) vectors are derived by summing the word vectors
  - Not a real semantic composition
    - *The man broke the vase* vs. *The vase broke the man*
- Cf. Kintsch (2001) for a more sophisticated LSA-based approach to predication
Beyond word meaning

- Few attempts to use WSMs to represent sentence meaning
- “Bag of words” approach (cf. Landauer and Dumais 1997)
  - Sentence (discourse) vectors are derived by summing the word vectors
  - Not a real semantic composition
    - The man broke the vase vs. The vase broke the man
- Cf. Kintsch (2001) for a more sophisticated LSA-based approach to predication
WSMs and compositional semantics

- Syntagmatic semantic properties represent crucial aspects of meaning and a stronghold of formal, symbolic models:
  - argument structure
  - selectional preferences
  - thematic roles, etc.

- Crucial issues:
  - To what extent can these properties be derived from word co-occurrences?
  - How can they be captured by WSMs?
WSMs and compositional semantics

- Syntagmatic semantic properties represent crucial aspects of meaning and a stronghold of formal, symbolic models:
  - argument structure
  - selectional preferences
  - thematic roles, etc.

- Crucial issues:
  - *To what extent can these properties be derived from word co-occurrences?*
  - *How can they be captured by WSMs?*
Conclusions

- Compared to other styles of meaning representation, WSMs are cognitively appealing
  - Avoid the problem of specifying *a priori* the semantic metalanguage (e.g. features, relations, etc.)
  - Word semantic properties can be derived from (fairly) simple input processing
  - Quite high predictive power in semantic behavioral tasks (cf. Vigliocco *et al.* 2004)
    - WSMs (LSA) > WordNet
    - *but...* speaker-generated features > WSMs (LSA)

- Still, many open issues on:
  - the design and structure of semantic representations based on word co-occurrences
  - the role of context features in shaping meaning
  - word space dynamics (e.g. learning, change, etc.)
Conclusions

- Compared to other styles of meaning representation, WSMs are cognitively appealing
  - Avoid the problem of specifying *a priori* the semantic metalanguage (e.g. features, relations, etc.)
  - Word semantic properties can be derived from (fairly) simple input processing
  - Quite high predictive power in semantic behavioral tasks (cf. Vigliocco *et al.* 2004)
    - WSMs (LSA) > WordNet
    - *but...* speaker-generated features > WSMs (LSA)
- Still, many open issues on:
  - the design and structure of semantic representations based on word co-occurrences
  - the role of context features in shaping meaning
  - word space dynamics (e.g. learning, change, etc.)
Conclusions

- Compared to other styles of meaning representation, WSMs are cognitively appealing
  - Avoid the problem of specifying *a priori* the semantic metalanguage (e.g. features, relations, etc.)
  - Word semantic properties can be derived from (fairly) simple input processing
  - Quite high predictive power in semantic behavioral tasks (cf. Vigliocco *et al.* 2004)
    - WSMs (LSA) > WordNet
    - *but...* speaker-generated features > WSMs (LSA)

- Still, many open issues on:
  - the design and structure of semantic representations based on word co-occurrences
  - the role of context features in shaping meaning
  - word space dynamics (e.g. learning, change, etc.)
Some references


