Hyperdimensional computing for human data meets the squinting linguist

Jussi Karlgren

Gavagai | KTH | (Currently Visiting Scholar at Stanford)

may 2018

This work is supported by Vinnova, the Swedish Governmental Agency for Innovation Systems under a Vinnmer Marie Curie Grant.

Jussi Karlgren

- stylistics, genre, statistical models of language in use, distributional semantics
- founder and researcher at the text analysis company Gavagai
- adjoint professor of language technology at KTH Royal Institute of Technology, Stockholm
- (visitor at Stanford University Dept of Linguistics, 2017-18)

Jussi Karlgren

- stylistics, genre, statistical models of language in use, distributional semantics
- founder and researcher at the text analysis company Gavagai
- adjoint professor of language technology at KTH Royal Institute of Technology, Stockholm
- (visitor at Stanford University Dept of Linguistics, 2017-18)

Jussi Karlgren

- stylistics, genre, statistical models of language in use, distributional semantics
- founder and researcher at the text analysis company Gavagai
- adjoint professor of language technology at KTH Royal Institute of Technology, Stockholm
- (visitor at Stanford University Dept of Linguistics, 2017-18)

Jussi Karlgren

- stylistics, genre, statistical models of language in use, distributional semantics
- founder and researcher at the text analysis company Gavagai
- adjoint professor of language technology at KTH Royal Institute of Technology, Stockholm
- (visitor at Stanford University Dept of Linguistics, 2017-18)

Gavagai

- Based in Stockholm (with roots at SICS, Swedish Institute for Computer Science)
- Research in processing human language since late 1990's
- Now a text analysis company with services in 45 languages, processing millions of documents daily
  - questionnaire processing explorer.gavagai.se
  - media monitoring monitor.gavagai.se
  - Gavagai living lexicon lexicon.gavagai.se

- Magnus Sahlgren et al. 2017. Gavagai Living Lexicon. LREC.
- Fredrik Espinoza et al. 2018. "Analysis of Open Answers to Survey Questions through Interactive Clustering and Theme Extraction." ACM CHIIR.
- Jussi Karlgren et al. 2012. "Usefulness of sentiment analysis." ECIR
- Magnus Sahlgren and Jussi Karlgren. 2009. "Terminology mining in social media." CIKM
- Pentti Kanerva et al. 2001. "Computing with large random patterns." CSLI

Gavagai

- Based in Stockholm (with roots at SICS, Swedish Institute for Computer Science)
- Research in processing human language since late 1990's
- Now a text analysis company with services in 45 languages, processing millions of documents daily
  - questionnaire processing explorer.gavagai.se
  - media monitoring monitor.gavagai.se
  - Gavagai living lexicon lexicon.gavagai.se

- Magnus Sahlgren et al. 2017. Gavagai Living Lexicon. LREC.
- Fredrik Espinoza et al. 2018. "Analysis of Open Answers to Survey Questions through Interactive Clustering and Theme Extraction." ACM CHIIR.
- Jussi Karlgren et al. 2012. "Usefulness of sentiment analysis." ECIR
- Magnus Sahlgren and Jussi Karlgren. 2009. "Terminology mining in social media." CIKM
- Pentti Kanerva et al. 2001. "Computing with large random patterns." CSLI

Gavagai

- Based in Stockholm (with roots at SICS, Swedish Institute for Computer Science)
- Research in processing human language since late 1990's
- Now a text analysis company with services in 45 languages, processing millions of documents daily
  - questionnaire processing explorer.gavagai.se
  - media monitoring monitor.gavagai.se
  - Gavagai living lexicon lexicon.gavagai.se

- Magnus Sahlgren et al. 2017. Gavagai Living Lexicon. LREC.
- Fredrik Espinoza et al. 2018. "Analysis of Open Answers to Survey Questions through Interactive Clustering and Theme Extraction." ACM CHIIR.
- Jussi Karlgren et al. 2012. "Usefulness of sentiment analysis." ECIR
- Magnus Sahlgren and Jussi Karlgren. 2009. "Terminology mining in social media." CIKM
- Pentti Kanerva et al. 2001. "Computing with large random patterns." CSLI

Gavagai

- Based in Stockholm (with roots at SICS, Swedish Institute for Computer Science)
- Research in processing human language since late 1990's
- Now a text analysis company with services in 45 languages, processing millions of documents daily
  - questionnaire processing explorer.gavagai.se
  - media monitoring monitor.gavagai.se
  - Gavagai living lexicon lexicon.gavagai.se

- Magnus Sahlgren et al. 2017. Gavagai Living Lexicon. LREC.
- Fredrik Espinoza et al. 2018. "Analysis of Open Answers to Survey Questions through Interactive Clustering and Theme Extraction." ACM CHIIR.
- Jussi Karlgren et al. 2012. "Usefulness of sentiment analysis." ECIR
- Magnus Sahlgren and Jussi Karlgren. 2009. "Terminology mining in social media." CIKM
- Pentti Kanerva et al. 2001. "Computing with large random patterns." CSLI

Gavagai

- Based in Stockholm (with roots at SICS, Swedish Institute for Computer Science)
- Research in processing human language since late 1990's
- Now a text analysis company with services in 45 languages, processing millions of documents daily
  - questionnaire processing explorer.gavagai.se
  - media monitoring monitor.gavagai.se
  - Gavagai living lexicon lexicon.gavagai.se

- Magnus Sahlgren et al. 2017. Gavagai Living Lexicon. LREC.
- Fredrik Espinoza et al. 2018. "Analysis of Open Answers to Survey Questions through Interactive Clustering and Theme Extraction." ACM CHIIR.
- Jussi Karlgren et al. 2012. "Usefulness of sentiment analysis." ECIR
- Magnus Sahlgren and Jussi Karlgren. 2009. "Terminology mining in social media." CIKM
- Pentti Kanerva et al. 2001. "Computing with large random patterns." CSLI

Gavagai

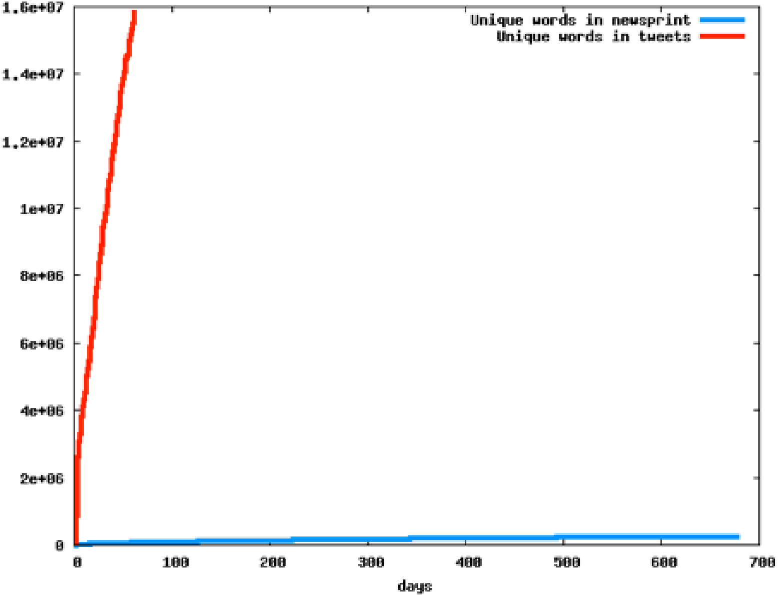
- Based in Stockholm (with roots at SICS, Swedish Institute for Computer Science)
- Research in processing human language since late 1990's
- Now a text analysis company with services in 45 languages, processing millions of documents daily
    - questionnaire processing explorer.gavagai.se
    - media monitoring monitor.gavagai.se
    - Gavagai living lexicon lexicon.gavagai.se

- Magnus Sahlgren et al. 2017. Gavagai Living Lexicon. LREC.
- Fredrik Espinoza et al. 2018. "Analysis of Open Answers to Survey Questions through Interactive Clustering and Theme Extraction." ACM CHIIR.
- Jussi Karlgren et al. 2012. "Usefulness of sentiment analysis." ECIR
- Magnus Sahlgren and Jussi Karlgren. 2009. "Terminology mining in social media." CIKM
- Pentti Kanerva et al. 2001. "Computing with large random patterns." CSLI

the pattern is more interesting than the data points

- new services, e.g. Internet of Things, sensor networks, human-generated data, universal logging: streaming data from many devices, many people, many levels of abstraction

language is a pilot case for high volume and high variety data



human information processing

- human information processing is effective for streaming data
- handles analogy & saliency
- observes patterns and change rather than the literal
- operates with self-learning rather than instruction

human information processing

- human information processing is effective for streaming data
- handles analogy & saliency
- observes patterns and change rather than the literal
- operates with self-learning rather than instruction

human information processing

- human information processing is effective for streaming data
- handles analogy & saliency
- observes patterns and change rather than the literal
- operates with self-learning rather than instruction

human information processing

- human information processing is effective for streaming data
- handles analogy & saliency
- observes patterns and change rather than the literal
- operates with self-learning rather than instruction

human information processing

- human information processing is effective for streaming data
- handles analogy & saliency
- observes patterns and change rather than the literal
- operates with self-learning rather than instruction

worth keeping in mind as a model

requirements for a knowledge representation

a representation should:

- have descriptive and explanatory power

  (allow backtracking into observations)
- be practical and convenient for further application

  (retain feature structure)
- be reasonably true to human performance

  (handle streaming and analogy!)
- handle patchy data

  (provide support for generalisation, defaults and constraints)
- be computationally habitable

  (not grow superlinearly with data input)
- be general

  (not tightly bound to some task)

requirements for a knowledge representation

a representation should:

- have descriptive and explanatory power

  (allow backtracking into observations)
- be practical and convenient for further application

  (retain feature structure)
- be reasonably true to human performance

  (handle streaming and analogy!)
- handle patchy data

  (provide support for generalisation, defaults and constraints)
- be computationally habitable

  (not grow superlinearly with data input)
- be general

  (not tightly bound to some task)

requirements for a knowledge representation

a representation should:

- have descriptive and explanatory power

  (allow backtracking into observations)
- be practical and convenient for further application

  (retain feature structure)
- be reasonably true to human performance

  (handle streaming and analogy!)
- handle patchy data

  (provide support for generalisation, defaults and constraints)
- be computationally habitable

  (not grow superlinearly with data input)
- be general

  (not tightly bound to some task)

requirements for a knowledge representation

a representation should:

- have descriptive and explanatory power

  (allow backtracking into observations)
- be practical and convenient for further application

  (retain feature structure)
- be reasonably true to human performance

  (handle streaming and analogy!)
- handle patchy data

  (provide support for generalisation, defaults and constraints)
- be computationally habitable

  (not grow superlinearly with data input)
- be general

  (not tightly bound to some task)

requirements for a knowledge representation

a representation should:

- have descriptive and explanatory power

  (allow backtracking into observations)
- be practical and convenient for further application

  (retain feature structure)
- be reasonably true to human performance

  (handle streaming and analogy!)
- handle patchy data

  (provide support for generalisation, defaults and constraints)
- be computationally habitable

  (not grow superlinearly with data input)
- be general

  (not tightly bound to some task)

requirements for a knowledge representation

a representation should:

- have descriptive and explanatory power

  (allow backtracking into observations)
- be practical and convenient for further application

  (retain feature structure)
- be reasonably true to human performance

  (handle streaming and analogy!)
- handle patchy data

  (provide support for generalisation, defaults and constraints)
- be computationally habitable

  (not grow superlinearly with data input)
- be general

  (not tightly bound to some task)

requirements for a knowledge representation

a representation should:

- have descriptive and explanatory power

  (allow backtracking into observations)
- be practical and convenient for further application

  (retain feature structure)
- be reasonably true to human performance

  (handle streaming and analogy!)
- handle patchy data

  (provide support for generalisation, defaults and constraints)
- be computationally habitable

  (not grow superlinearly with data input)
- be general

  (not tightly bound to some task)

high-dimensional computing

the approach suggested by us is

- high-dimensional

  to allow a rich representation
- implemented as a vector space

  mathematically well defined and manageable for implementation
- uses random patterns to index observations

  achieves orthogonality for all practical purposes

- Pentti Kanerva. 2009. "Hyperdimensional computing: An introduction to computing in distributed representation with high-dimensional random vectors." Cognitive Computation 1:2.

high-dimensional computing

the approach suggested by us is

- high-dimensional

  to allow a rich representation

- implemented as a vector space

  mathematically well defined and manageable for implementation

- uses random patterns to index observations

  achieves orthogonality for all practical purposes

- Pentti Kanerva. 2009. "Hyperdimensional computing: An introduction to computing in distributed representation with high-dimensional random vectors." Cognitive Computation 1:2.

high-dimensional computing

the approach suggested by us is

- high-dimensional

  to allow a rich representation

- implemented as a vector space

  mathematically well defined and manageable for implementation

- uses random patterns to index observations

  achieves orthogonality for all practical purposes

- Pentti Kanerva. 2009. "Hyperdimensional computing: An introduction to computing in distributed representation with high-dimensional random vectors." Cognitive Computation 1:2.

high-dimensional computing

the approach suggested by us is

- high-dimensional

  to allow a rich representation

- implemented as a vector space

  mathematically well defined and manageable for implementation

- uses random patterns to index observations

  achieves orthogonality for all practical purposes

- Pentti Kanerva. 2009. "Hyperdimensional computing: An introduction to computing in distributed representation with high-dimensional random vectors." Cognitive Computation 1:2.

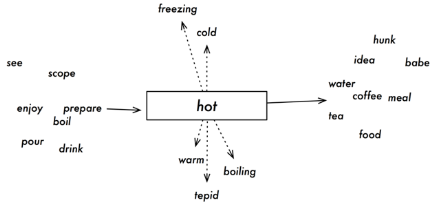
distributional semantics

the weather is great in barcelona
the weather is gray in stockholm
the weather is hot in lodi
the climate is passable in nice
the weather is chilly in helsinki
the weather is nippy in moscow
the weather is nice in hong kong
the weather in syktyvkar is balmy
the climate is chilly at the office
the tea is hot
i drink tea
a hot meal will make you feel better
enjoy your hot beverages

- Zellig Harris. 1968. Mathematical structures of language.

distributional semantics

the weather is great in barcelona
the weather is gray in stockholm
the weather is hot in lodi
the climate is passable in nice
the weath
the weath
the weath
the weath
the clima
the tea i
i drink tea
a hot meal will make you feel better
enjoy your hot beverages



- Zellig Harris. 1968. Mathematical structures of language.

implementational model: embeddings

- (relatively) low-dimensional projections of high-dimensional observational data and their cooccurrence
- generalises over cooccurrence events
- quantifies distributional data
- (surprisingly) not (often) systematically used in broader data contexts
- risk of low explainability

hyperdimensional computing in practice

- random index vectors or labels for basic features
- aggregated context vectors for cooccurrences of features
- similarity between vectors can be measured by cosine
- operations on vectors:
  - addition which yields a similar vector to the inputs
  - multiplication which yields a dissimilar vector to its inputs but preserves relative similarity
  - permutation allows for e.g. sequences or tensors to be represented in one vector
- allows for explicit feature engineering if necessary

hyperdimensional computing in practice

- random index vectors or labels for basic features
- aggregated context vectors for cooccurrences of features
- similarity between vectors can be measured by cosine
- operations on vectors:
  - addition which yields a similar vector to the inputs
  - multiplication which yields a dissimilar vector to its inputs but preserves relative similarity
  - permutation allows for e.g. sequences or tensors to be represented in one vector
- allows for explicit feature engineering if necessary

hyperdimensional computing in practice

- random index vectors or labels for basic features
- aggregated context vectors for cooccurrences of features
- similarity between vectors can be measured by cosine
- operations on vectors:
  - addition which yields a similar vector to the inputs
  - multiplication which yields a dissimilar vector to its inputs but preserves relative similarity
  - permutation allows for e.g. sequences or tensors to be represented in one vector
- allows for explicit feature engineering if necessary

hyperdimensional computing in practice

- random index vectors or labels for basic features
- aggregated context vectors for cooccurrences of features
- similarity between vectors can be measured by cosine
- operations on vectors:
  - addition which yields a similar vector to the inputs
  - multiplication which yields a dissimilar vector to its inputs but preserves relative similarity
  - permutation allows for e.g. sequences or tensors to be represented in one vector
- allows for explicit feature engineering if necessary

hyperdimensional computing in practice

- random index vectors or labels for basic features
- aggregated context vectors for cooccurrences of features
- similarity between vectors can be measured by cosine
- operations on vectors:
  - addition which yields a similar vector to the inputs
  - multiplication which yields a dissimilar vector to its inputs but preserves relative similarity
  - permutation allows for e.g. sequences or tensors to be represented in one vector
- allows for explicit feature engineering if necessary

hyperdimensional computing in practice

- random index vectors or labels for basic features
- aggregated context vectors for cooccurrences of features
- similarity between vectors can be measured by cosine
- operations on vectors:
  - addition which yields a similar vector to the inputs
  - multiplication which yields a dissimilar vector to its inputs but preserves relative similarity
  - permutation allows for e.g. sequences or tensors to be represented in one vector
- allows for explicit feature engineering if necessary

hyperdimensional computing in practice

- random index vectors or labels for basic features
- aggregated context vectors for cooccurrences of features
- similarity between vectors can be measured by cosine
- operations on vectors:
  - addition which yields a similar vector to the inputs
  - multiplication which yields a dissimilar vector to its inputs but preserves relative similarity
  - permutation allows for e.g. sequences or tensors to be represented in one vector
- allows for explicit feature engineering if necessary

hyperdimensional computing in practice

- random index vectors or labels for basic features
- aggregated context vectors for cooccurrences of features
- similarity between vectors can be measured by cosine
- operations on vectors:
  - addition which yields a similar vector to the inputs
  - multiplication which yields a dissimilar vector to its inputs but preserves relative similarity
  - permutation allows for e.g. sequences or tensors to be represented in one vector
- allows for explicit feature engineering if necessary

```
   hot  [0, 0, 0, 0, 0, 0]
chilly  [0, 0, 0, 0, 0, 0]
```

```
the weather is hot ... the weather is chilly ...
```

```
weather: [ 0, -1, 0, 1, 0, 0]
```

```
   hot  [0, -1, 0, 1, 0, 0]
chilly  [0, -1, 0, 1, 0, 0]
```

the weather is hot … the weather is chilly … the climate is hot … the climate is chilly …

```
weather: [ 0, -1, 0, 1, 0, 0]
climate: [ 1, 0, -1, 0, 0, 0]
```

```
   hot  [1, -1, -1, 1, 0, 0]
chilly  [1, -1, -1, 1, 0, 0]
```

the weather is hot … the weather is chilly … the climate is hot … the climate is chilly …

the weather turned hot … the climate turned chilly …

weather: [0, -1, 0, 1, 0, 0]
climate: [1, 0, -1, 0, 0, 0]

hot	[2, -1, -2, 1, 0, 0]
chilly	[1, -2, -1, 2, 0, 0]

the weather is hot … the weather is chilly … the climate is hot … the climate is chilly …

the weather turned hot … the climate turned chilly …

weather: [0, -1, 0, 1, 0, 0]
climate: [1, 0, -1, 0, 0, 0]

hot [2, -1, -2, 1, 0, 0]
chilly [1, -2, -1, 2, 0, 0]

many parameters, but simple experimentation!

- Pentti Kanerva, Jan Kristoferson, and Anders Holst. 2000. Random Indexing of Text Samples for Latent Semantic Analysis. CogSci.
- Magnus Sahlgren, Anders Holst, and Pentti Kanerva. 2008. Permutations as a Means to Encode Order in Word Space. CogSci.

quantitative characteristics

knowledge representation: features for observations, aggregated into states, allow verification, decomposition, and various calculus operations

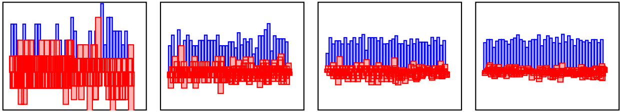
a holographic model using random patterns allows vastly larger feature palette given a preset dimensionality of representation

noise loss, twenty items vs random correlation, in 100, 500, 1000, 2000 dimensions

quantitative characteristics

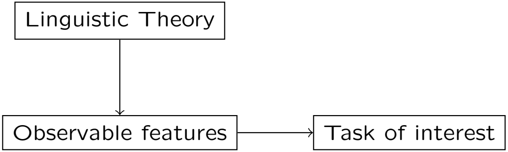
knowledge representation: features for observations, aggregated into states, allow verification, decomposition, and various calculus operations

a holographic model using random patterns allows vastly larger feature palette given a preset dimensionality of representation



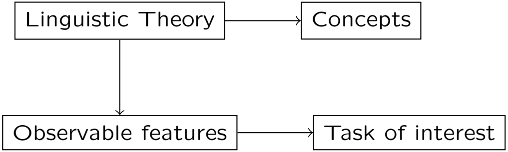
noise loss, twenty items vs random correlation, in 100, 500, 1000, 2000 dimensions

the squinting linguist

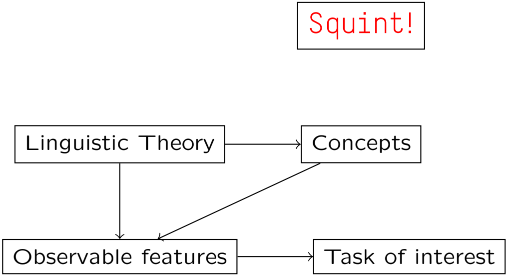


the squinting linguist

Squint!

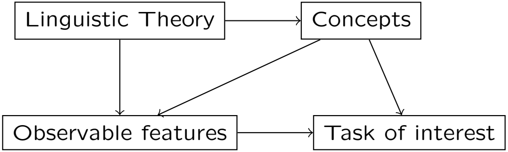


the squinting linguist



the squinting linguist

Squint!



loads of sentences, represented as sums of features

(2000 dimensions; features are 10 non-zero cells):

loads of sentences, represented as sums of features (2000 dimensions; features are 10 non-zero cells):

(a) words

loads of sentences, represented as sums of features (2000 dimensions; features are 10 non-zero cells):

(a) words

(b) constructional elements: negations, amplifiers, …

loads of sentences, represented as sums of features (2000 dimensions; features are 10 non-zero cells):

(a) words

(b) constructional elements: negations, amplifiers, …

find neighbours to:

„I really did not like the clarinet, I am afraid: it sounded weak!”

loads of sentences, represented as sums of features (2000 dimensions; features are 10 non-zero cells):

(a) words

(b) constructional elements: negations, amplifiers, …

find neighbours to:

„I really did not like the clarinet, I am afraid: it sounded weak!”

(a) words:
My sister plays the clarinet.

loads of sentences, represented as sums of features (2000 dimensions; features are 10 non-zero cells):

(a) words

(b) constructional elements: negations, amplifiers, …

find neighbours to:

„I really did not like the clarinet, I am afraid: it sounded weak!”

(a) words:
My sister plays the clarinet.

(b) constructions:
I’m surrounded by really soft decadent pillows which do not work for me at all.

loads of sentences, represented as sums of features (2000 dimensions; features are 10 non-zero cells):

(a) words

(b) constructional elements: negations, amplifiers, …

find neighbours to:

„I really did not like the clarinet, I am afraid: it sounded weak!"

(a) words:
My sister plays the clarinet.

(b) constructions:
I'm surrounded by really soft decadent pillows which do not work for me at all.

in the same representational space. great for hypothesis testing, and for differing tasks!

authorship gender profiling

linguistic theories are not built for this task and are overly specific;
word occurrence models overtrain on topic and do not generalise.

- https://pan.webis.de/clef18/pan18-web/author-profiling.html

authorship gender profiling

linguistic theories are not built for this task and are overly specific;
word occurrence models overtrain on topic and do not generalise.

- add all observed features of potential interest into same representation

---

- https://pan.webis.de/clef18/pan18-web/author-profiling.html

authorship gender profiling

linguistic theories are not built for this task and are overly specific;
word occurrence models overtrain on topic and do not generalise.

- add all observed features of potential interest into same representation
- use cosine to test which representations fit best

- https://pan.webis.de/clef18/pan18-web/author-profiling.html

authorship gender profiling

linguistic theories are not built for this task and are overly specific;
word occurrence models overtrain on topic and do not generalise.

- add all observed features of potential interest into same representation
- use cosine to test which representations fit best
- findings (so far):
  - ♀: first person subjects, „truly"-amplifiers
  - ♂: think verbs and hedges

- https://pan.webis.de/clef18/pan18-web/author-profiling.html

why not just use existing models?

- we believe this to be the only sustainable computational approach for knowledge representation in applications where data is streaming, varied, and at real-world scale
- at gavagai we have found no reason to depart from the basic premises for the purpose of processing human language
- this is in contrast to
  - localist representations,
  - compiling models,
  - and end-to-end black boxes.

why not just use existing models?

- we believe this to be the only sustainable computational approach for knowledge representation in applications where data is streaming, varied, and at real-world scale
- at gavagai we have found no reason to depart from the basic premises for the purpose of processing human language
- this is in contrast to
  - localist representations,
  - compiling models,
  - and end-to-end black boxes.

why not just use existing models?

- we believe this to be the only sustainable computational approach for knowledge representation in applications where data is streaming, varied, and at real-world scale
- at gavagai we have found no reason to depart from the basic premises for the purpose of processing human language
- this is in contrast to
  - localist representations,
  - compiling models,
  - and end-to-end black boxes.

why not just use existing models?

- we believe this to be the only sustainable computational approach for knowledge representation in applications where data is streaming, varied, and at real-world scale
- at gavagai we have found no reason to depart from the basic premises for the purpose of processing human language
- this is in contrast to
  - localist representations,
  - compiling models,
  - and end-to-end black boxes.

why not just use existing models?

- we believe this to be the only sustainable computational approach for knowledge representation in applications where data is streaming, varied, and at real-world scale
- at gavagai we have found no reason to depart from the basic premises for the purpose of processing human language
- this is in contrast to
  - localist representations,
  - compiling models,
  - and end-to-end black boxes.

why not just use existing models?

- we believe this to be the only sustainable computational approach for knowledge representation in applications where data is streaming, varied, and at real-world scale
- at gavagai we have found no reason to depart from the basic premises for the purpose of processing human language
- this is in contrast to
  - localist representations,
  - compiling models,
  - and end-to-end black boxes.

take home

1.
   - handing hypotheses and providing explanatory power are important ...
   - ... as is computational habitality ...
   - ... neither is optional.

take home

1. - handing hypotheses and providing explanatory power are important ...
   - ... as is computational habitality ...
   - ... neither is optional.
2. - feature engineering is a useful method to understand the world ...
   - ... knowledge representations for processing large amounts of data should support it.