Hierarchical Paragraph Vectors (HPV)

Extension to “Distributed Representations of Sentences and Documents”

Master Thesis Presentation

2015-09-21, Lukas Elmer

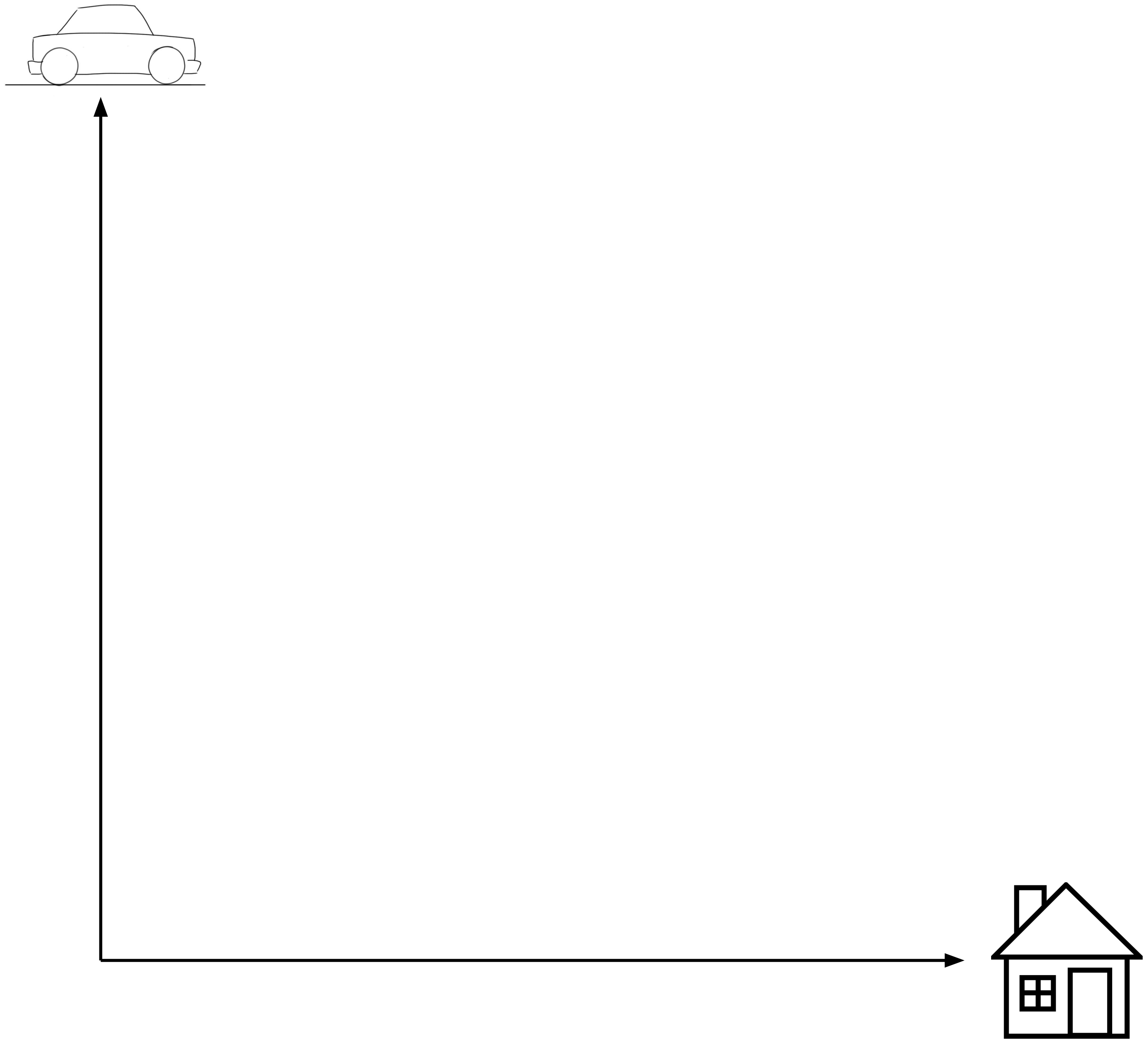
Overview

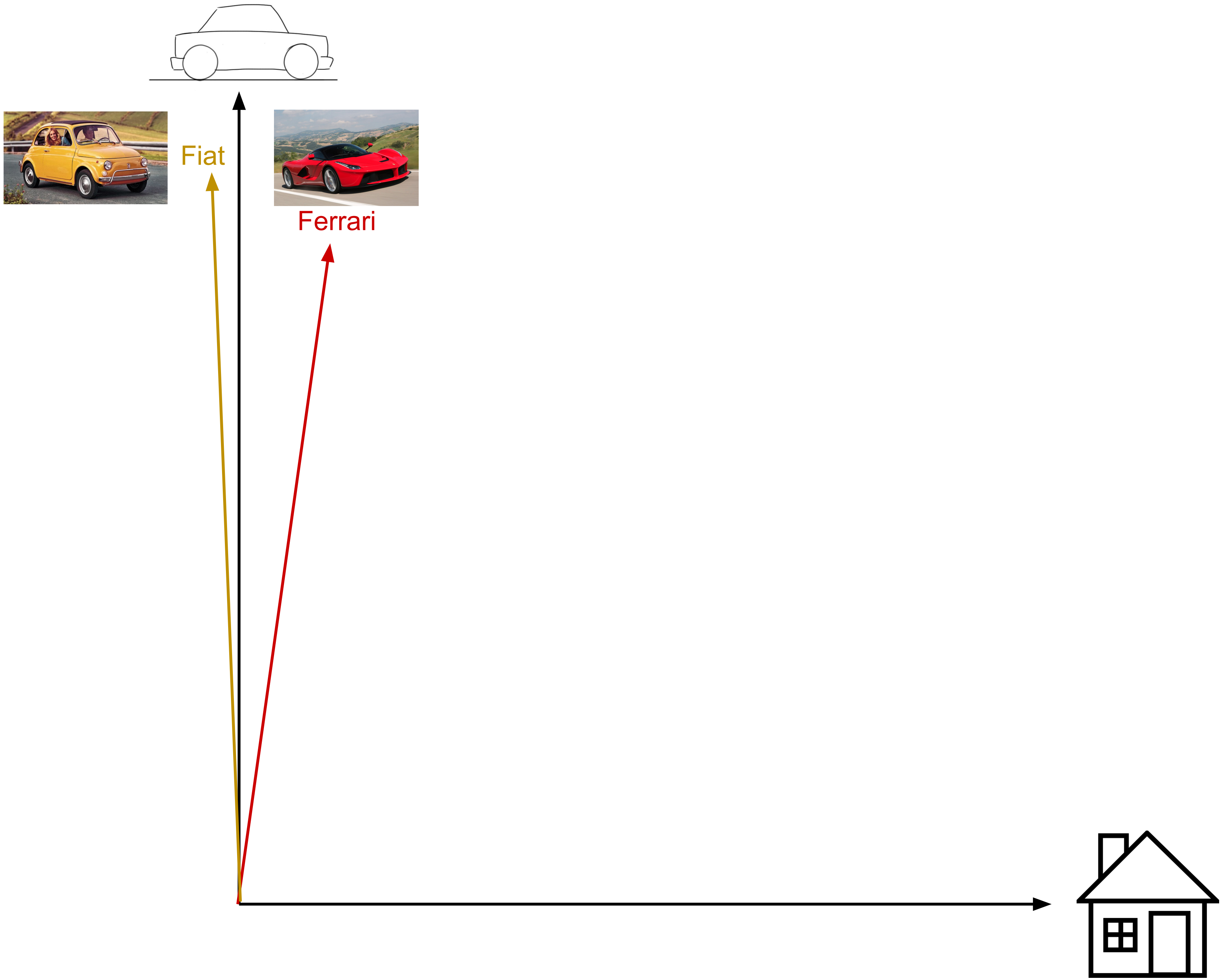
1) Foundations
2) Hierarchical paragraph vectors (HPV)
3) Things done
4) One real world problem
5) Results
6) Conclusion and future work

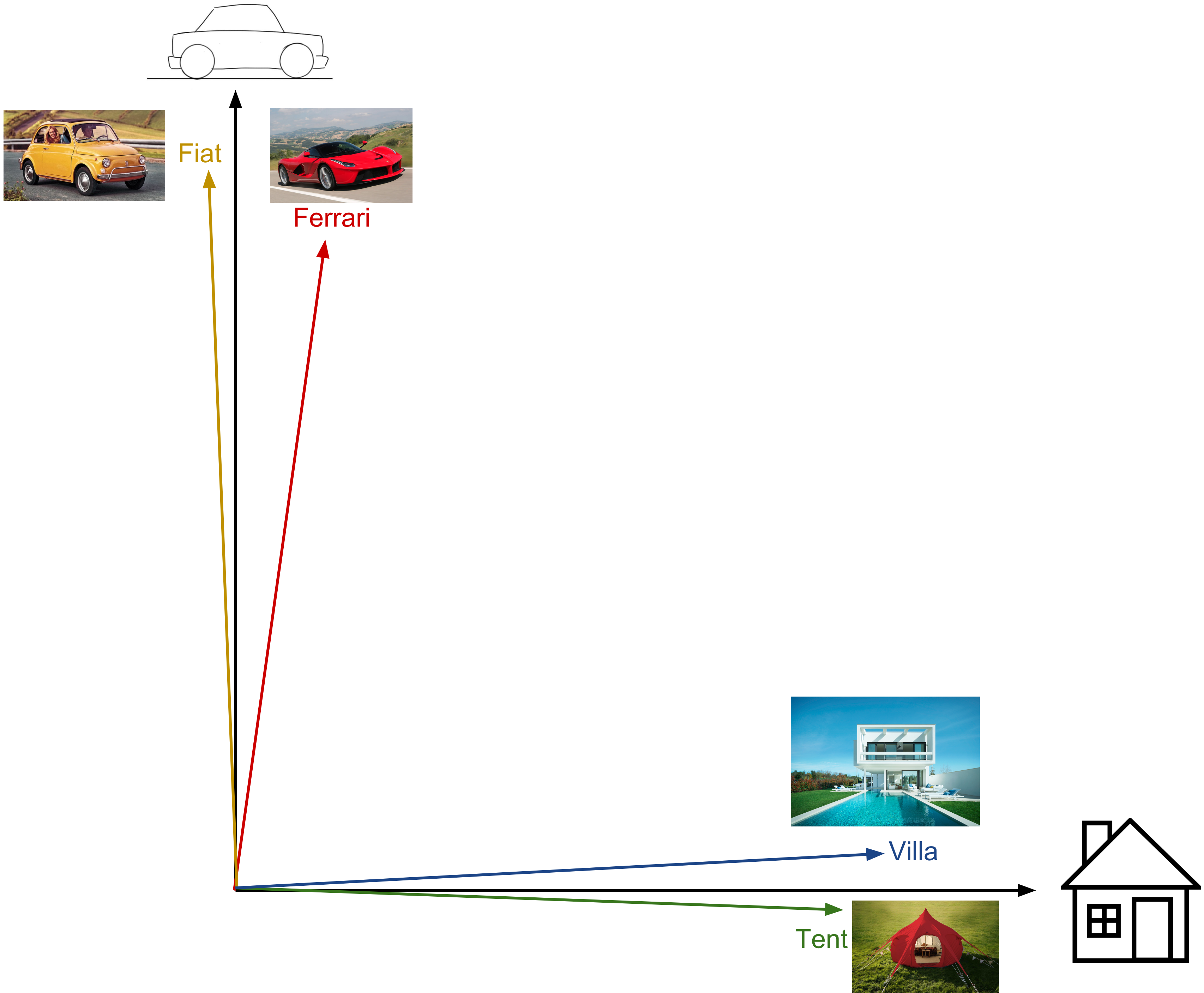
1) Foundations

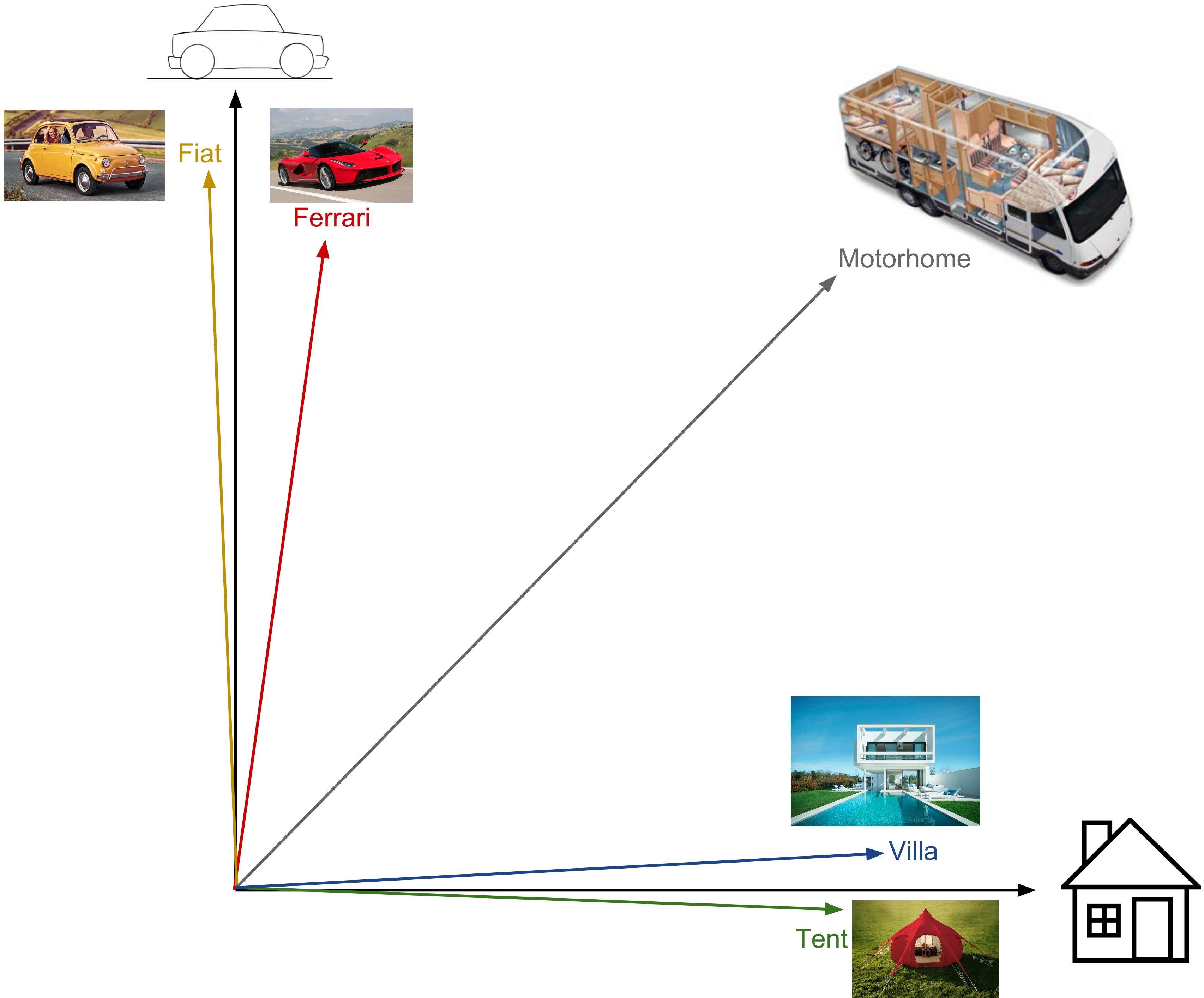
Primary Objective

- Represent *words* and *documents* as *vectors*
  - Fixed length
  - Low-dimensional
  - Dense
  - Useful

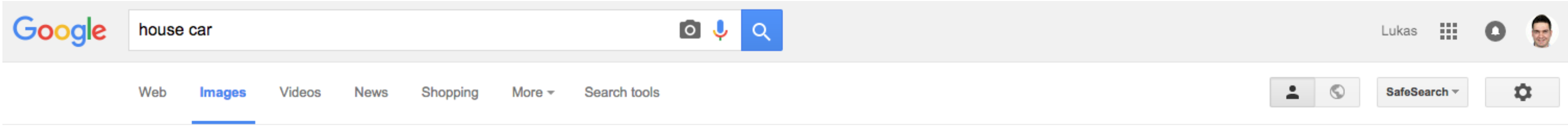


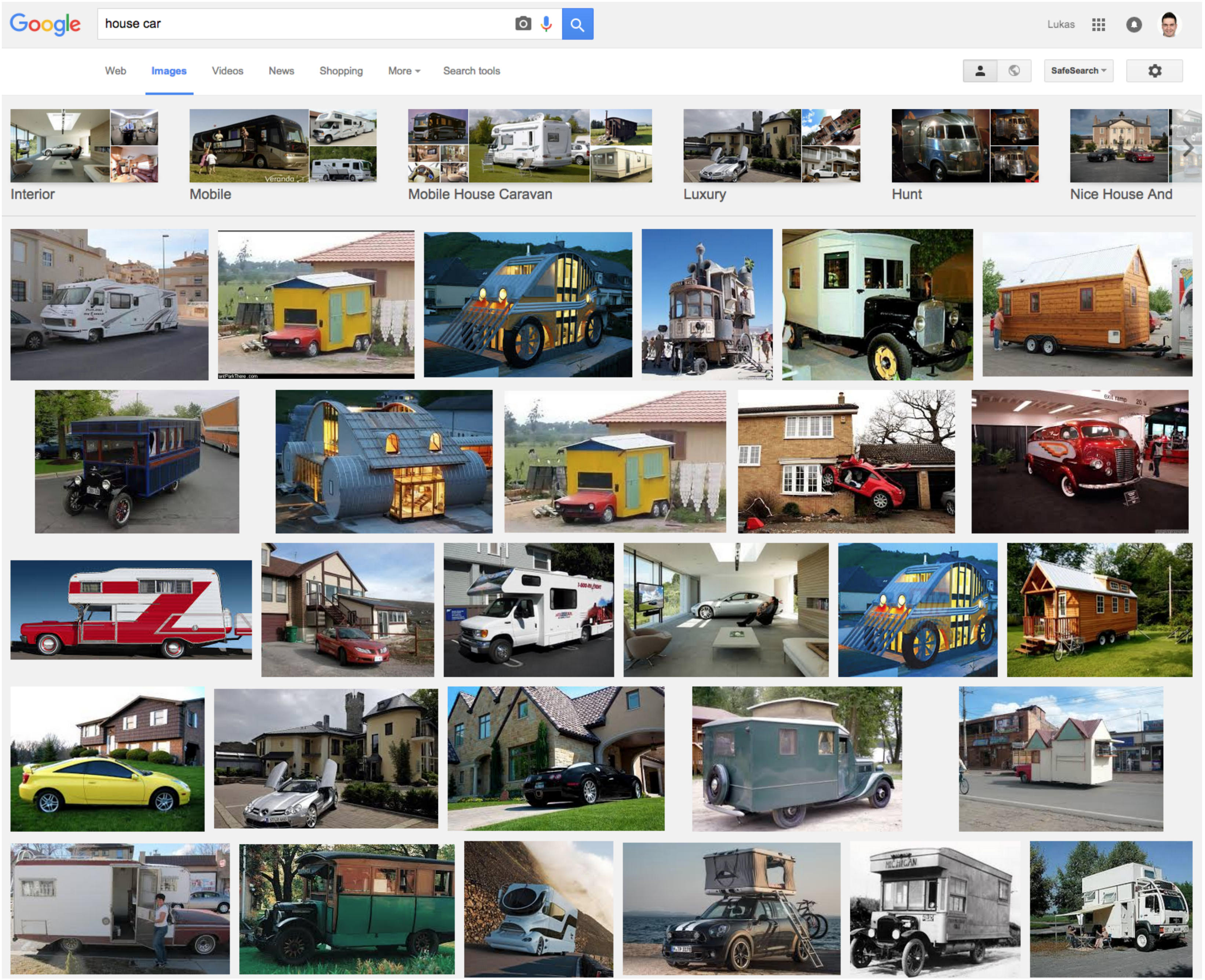

Fiat
Ferrari


Fiat
Ferrari
Villa
Tent



Fiat
Ferrari
Motorhome
Villa
Tent


Google
house car
Lukas
Web
Images
Videos
News
Shopping
More
Search tools
SafeSearch


Google
house car
Lukas
Web
Images
Videos
News
Shopping
More
Search tools
SafeSearch
Interior
Mobile
Mobile House Caravan
Luxury
Hunt
Nice House And

Standard Representations / Methods

Bag of Words

Document	Text
d_1	The cat sleeps on the sofa.
d_2	The mouse eats cheese.
d_3	The cat tries to catch the mouse.

Standard Representations / Methods

Bag of Words

Document	Text
d_1	The cat sleeps on the sofa.
d_2	The mouse eats cheese.
d_3	The cat tries to catch the mouse.

	the	cat	sleeps	on	sofa	mouse	eats	cheese	tries	to	catch
the	1	0	0	0	0	0	0	0	0	0	0
cat	0	1	0	0	0	0	0	0	0	0	0
...											
catch	0	0	0	0	0	0	0	0	0	0	1

Standard Representations / Methods

Bag of Words

	the	cat	sleeps	on	sofa	mouse	eats	cheese	tries	to	catch
the	1	0	0	0	0	0	0	0	0	0	0
cat	0	1	0	0	0	0	0	0	0	0	0
...											
catch	0	0	0	0	0	0	0	0	0	0	1

Document	Text
d_1	The cat sleeps on the sofa.
d_2	The mouse eats cheese.
d_3	The cat tries to catch the mouse.

Dimensionality, distance

Standard Representations / Methods

Bag of Words

Document	Text
d_1	The cat sleeps on the sofa.
d_2	The mouse eats cheese.
d_3	The cat tries to catch the mouse.

	the	cat	sleeps	on	sofa	mouse	eats	cheese	tries	to	catch
the	1	0	0	0	0	0	0	0	0	0	0
cat	0	1	0	0	0	0	0	0	0	0	0
...											
catch	0	0	0	0	0	0	0	0	0	0	1

	the	cat	sleeps	on	sofa	mouse	eats	cheese	tries	to	catch
d_1	2	1	1	1	1	0	0	0	0	0	0
d_2	1	0	0	0	0	1	1	1	0	0	0
d_3	2	1	0	0	0	1	0	0	1	1	1

Standard Representations / Methods

Bag of Words

Document	Text
d_1	The cat sleeps on the sofa.
d_2	The mouse eats cheese.
d_3	The cat tries to catch the mouse.

	the	cat	sleeps	on	sofa	mouse	eats	cheese	tries	to	catch
the	1	0	0	0	0	0	0	0	0	0	0
cat	0	1	0	0	0	0	0	0	0	0	0
...											
catch	0	0	0	0	0	0	0	0	0	0	1

	the	cat	sleeps	on	sofa	mouse	eats	cheese	tries	to	catch
d_1	2	1	1	1	1	0	0	0	0	0	0
d_2	1	0	0	0	0	1	1	1	0	0	0
d_3	2	1	0	0	0	1	0	0	1	1	1

Dimensionality, distance, word order

Standard Representations / Methods

Term Frequency – Inverse Document Frequency (TF–IDF)

Document	Text
d_1	The cat sleeps on the sofa.
d_2	The mouse eats cheese.
d_3	The cat tries to catch the mouse.

	cat	catch	cheese	eats	mouse	on	sleeps	sofa	the	to	tries
d_1	0.34	0.0	0.0	0.0	0.0	0.45	0.45	0.45	0.53	0.0	0.0
d_2	0.0	0.0	0.58	0.58	0.44	0.0	0.0	0.0	0.35	0.0	0.0
d_3	0.32	0.42	0.0	0.0	0.32	0.0	0.0	0.0	0.5	0.42	0.42

Standard Representations / Methods

Term Frequency – Inverse Document Frequency (TF–IDF)

Document	Text
d_1	The cat sleeps on the sofa.
d_2	The mouse eats cheese.
d_3	The cat tries to catch the mouse.

	cat	catch	cheese	eats	mouse	on	sleeps	sofa	the	to	tries
d_1	0.34	0.0	0.0	0.0	0.0	0.45	0.45	0.45	0.53	0.0	0.0
d_2	0.0	0.0	0.58	0.58	0.44	0.0	0.0	0.0	0.35	0.0	0.0
d_3	0.32	0.42	0.0	0.0	0.32	0.0	0.0	0.0	0.5	0.42	0.42

Dimensionality, distance, word order

Word2vec

Word2vec

Based on recurrent neural net language model (RNNLM), but more efficient

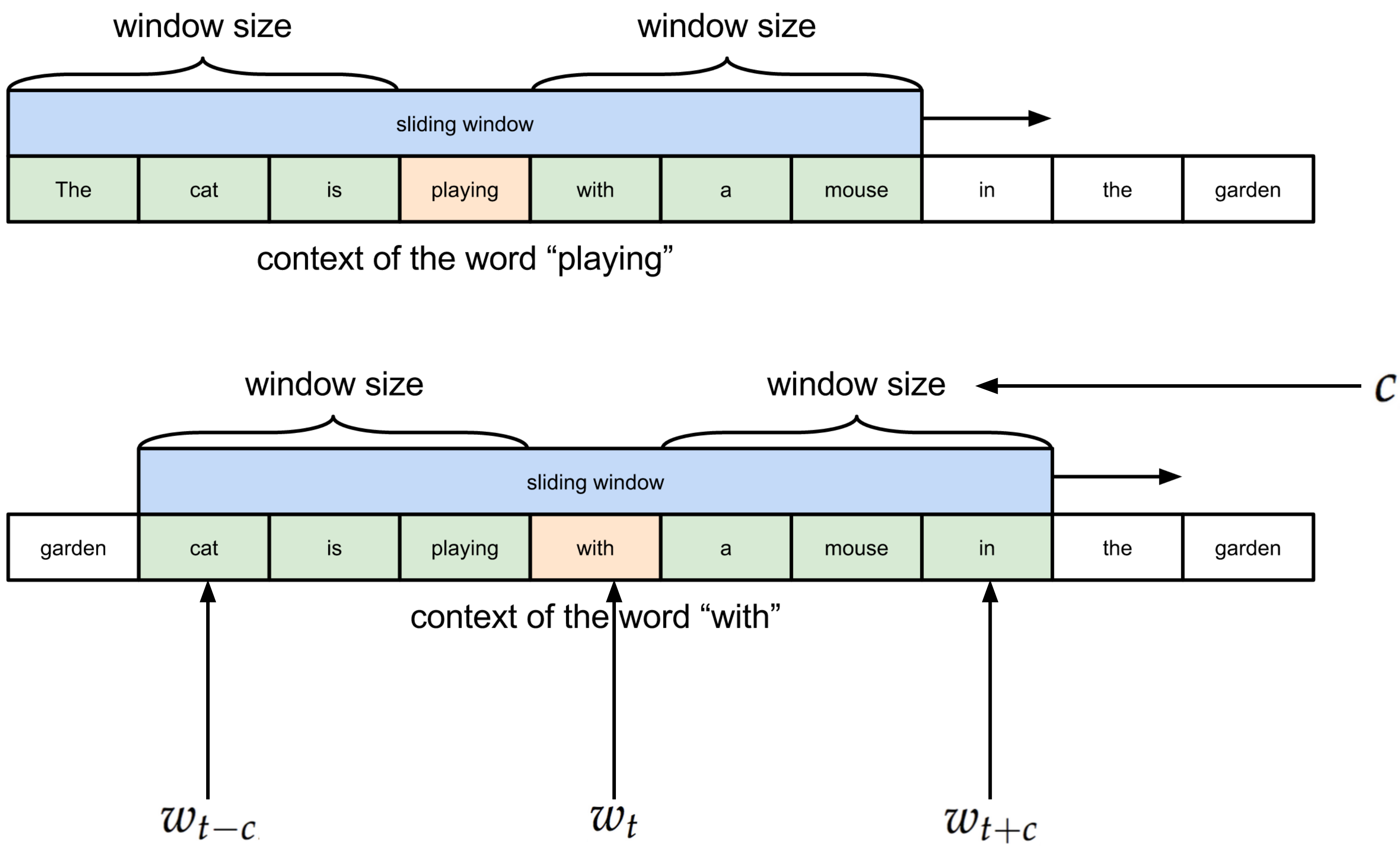
- NN with one hidden layer
- SGD
- Gradient obtained by backpropagation
- Predict conditional probabilities
  - (Hierarchical) Soft-max
  - Negative sampling

[1] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient Estimation of Word Representations in Vector Space. In Proceedings of Workshop at ICLR, 2013.

[2] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. Distributed Representations of Words and Phrases and their Compositionality. In Proceedings of NIPS, 2013.

[3] Tomas Mikolov, Wen-tau Yih, and Geoffrey Zweig. Linguistic Regularities in Continuous Space Word Representations. In Proceedings of NAACL HLT, 2013.

Word2vec Window Size and Context



Word2vec Objective

$$\underset{W}{\text{maximize}}\; p(w_t | w_{t-c},\ w_{t-(c-1)},\ \ldots,\ w_{t-1},\ w_{t+1},\ \ldots,\ w_{t+(c-1)},\ w_{t+c})$$

$$\underset{W}{\text{maximize}}\; \frac{1}{T} \sum_{t=1}^{T} \sum_{-c \le j \le c, j \neg 0} \log p(w_t | w_{t+j})$$

Word2vec Conditional Probabilities

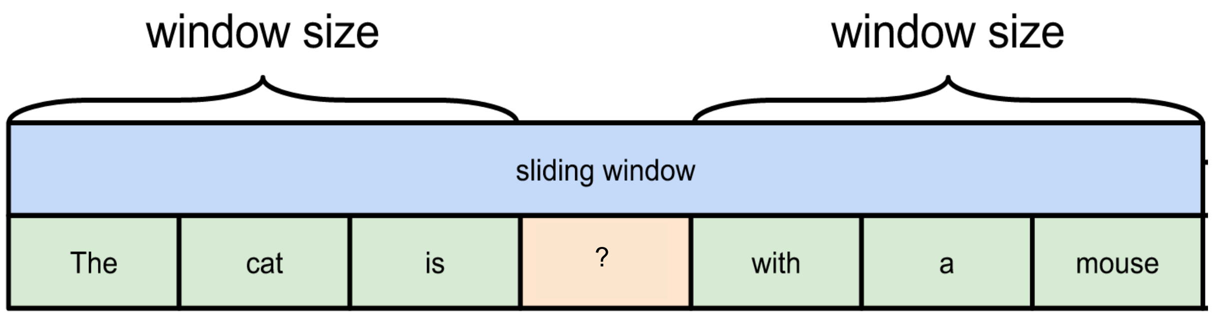
Soft-max $$p(w_t | w_{t-c}, \ldots, w_{t+c}) = \frac{e^{y_{w_t}}}{\sum_i e^{y_i}}$$

Optimizations

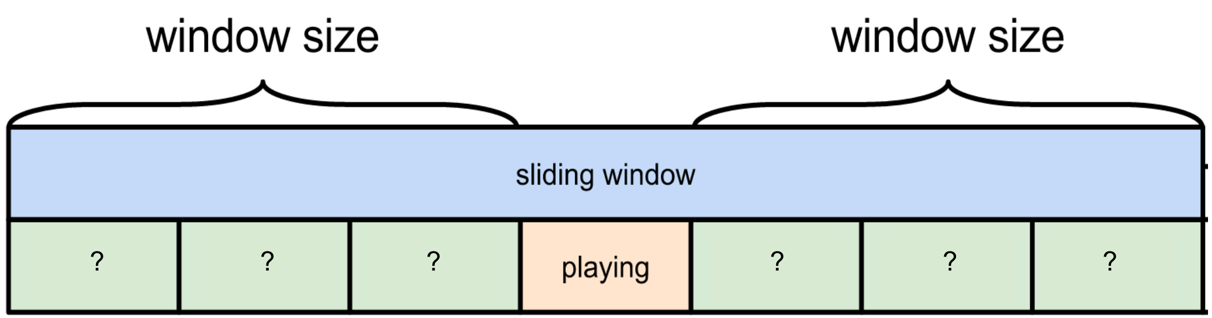
- Hierarchical soft-max
- Negative sampling

Word2vec Goal

CBOW

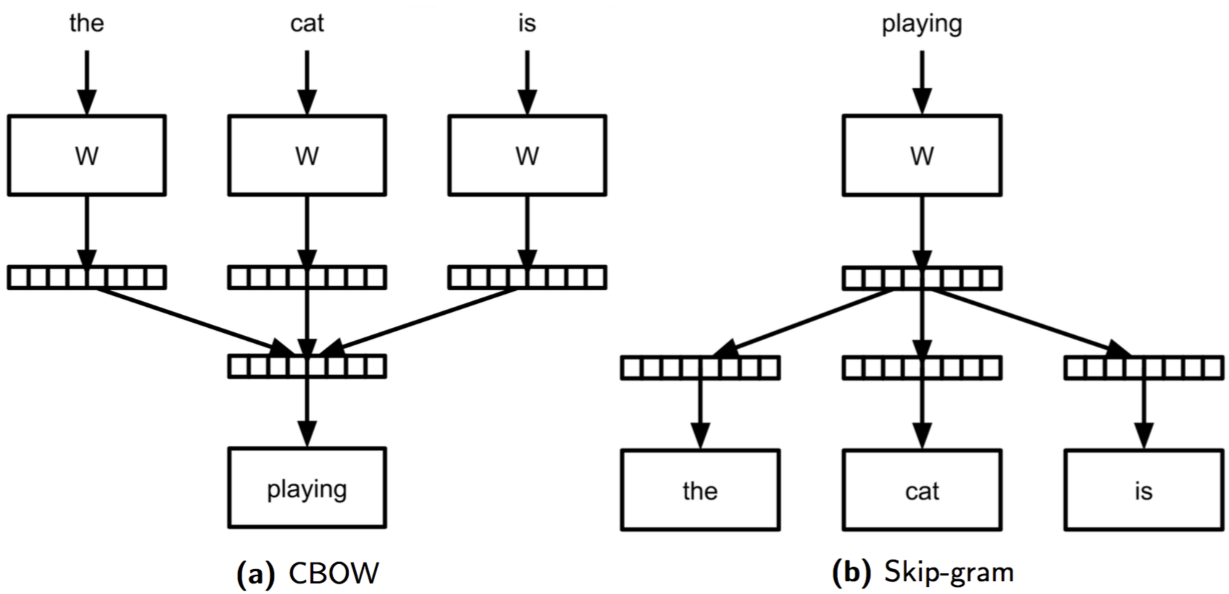


Skip-gram



Word2vec Goal

Word vectors help predicting



Doc2vec

Doc2vec

Paper

- Quoc V. Le and Tomas Mikolov. Distributed Representations of Sentences and Documents. CoRR, abs/1405.4, 2014.

Paragraph vector (PV)

Doc2vec - Paragraph Vectors PV

window size

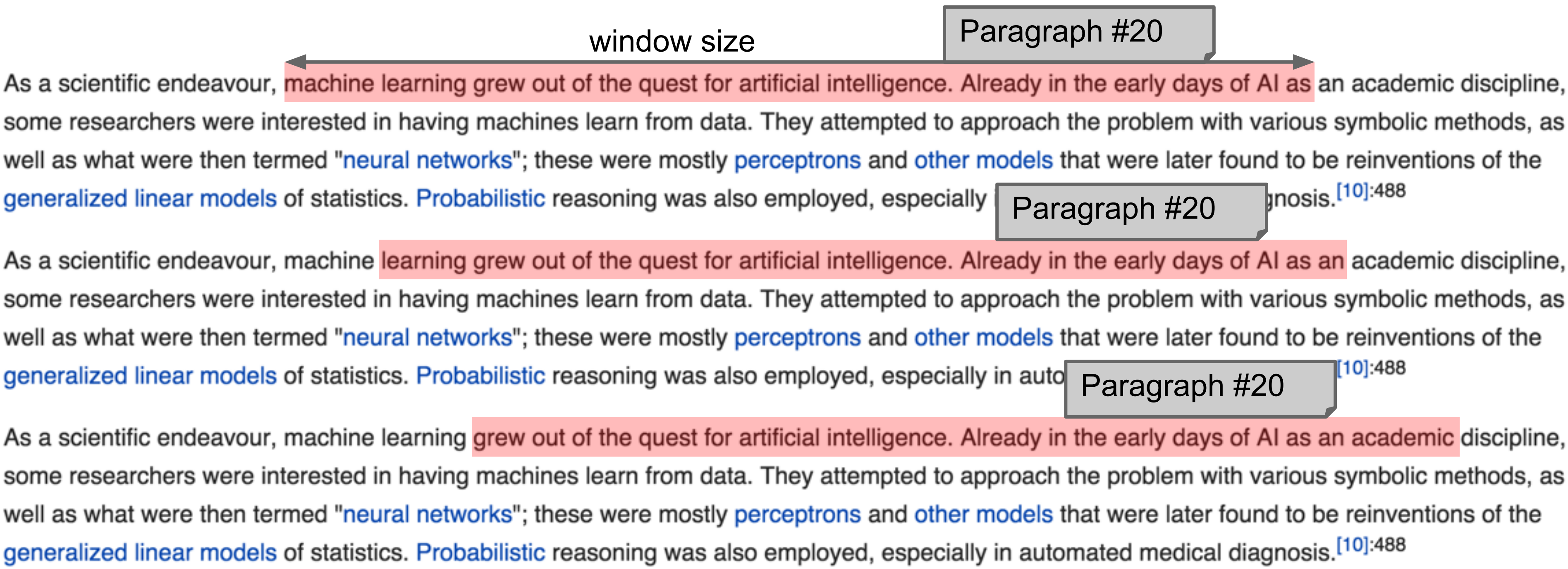
As a scientific endeavour, machine learning grew out of the quest for artificial intelligence. Already in the early days of AI as an academic discipline, some researchers were interested in having machines learn from data. They attempted to approach the problem with various symbolic methods, as well as what were then termed "neural networks"; these were mostly perceptrons and other models that were later found to be reinventions of the generalized linear models of statistics. Probabilistic reasoning was also employed, especially in automated medical diagnosis.[10]:488

As a scientific endeavour, machine learning grew out of the quest for artificial intelligence. Already in the early days of AI as an academic discipline, some researchers were interested in having machines learn from data. They attempted to approach the problem with various symbolic methods, as well as what were then termed "neural networks"; these were mostly perceptrons and other models that were later found to be reinventions of the generalized linear models of statistics. Probabilistic reasoning was also employed, especially in automated medical diagnosis.[10]:488

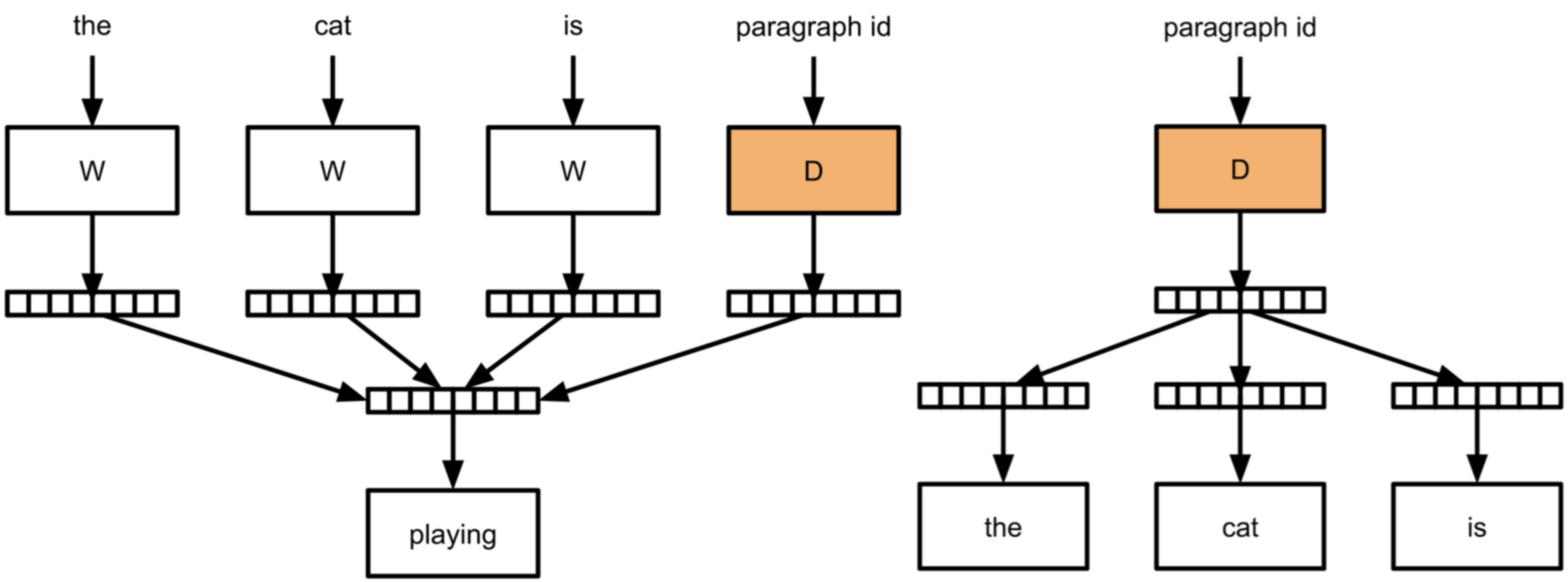
As a scientific endeavour, machine learning grew out of the quest for artificial intelligence. Already in the early days of AI as an academic discipline, some researchers were interested in having machines learn from data. They attempted to approach the problem with various symbolic methods, as well as what were then termed "neural networks"; these were mostly perceptrons and other models that were later found to be reinventions of the generalized linear models of statistics. Probabilistic reasoning was also employed, especially in automated medical diagnosis.[10]:488

Doc2vec - Paragraph Vectors PV

PV = memory for larger context



Doc2vec Goal



(a) PV-DM (b) PV-DBOW

2) Hierarchical Paragraph Vectors HPV

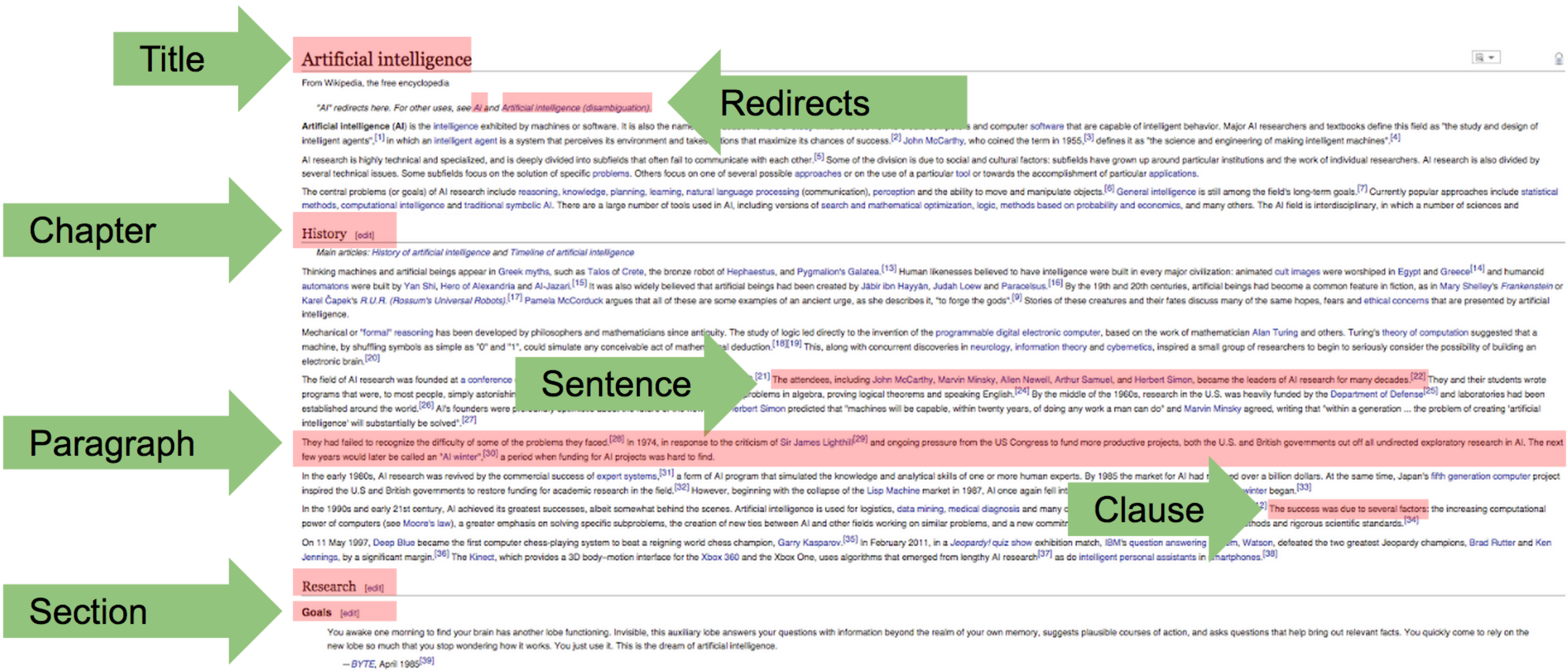
Hierarchies

Idea: have a hierarchy of paragraph vectors

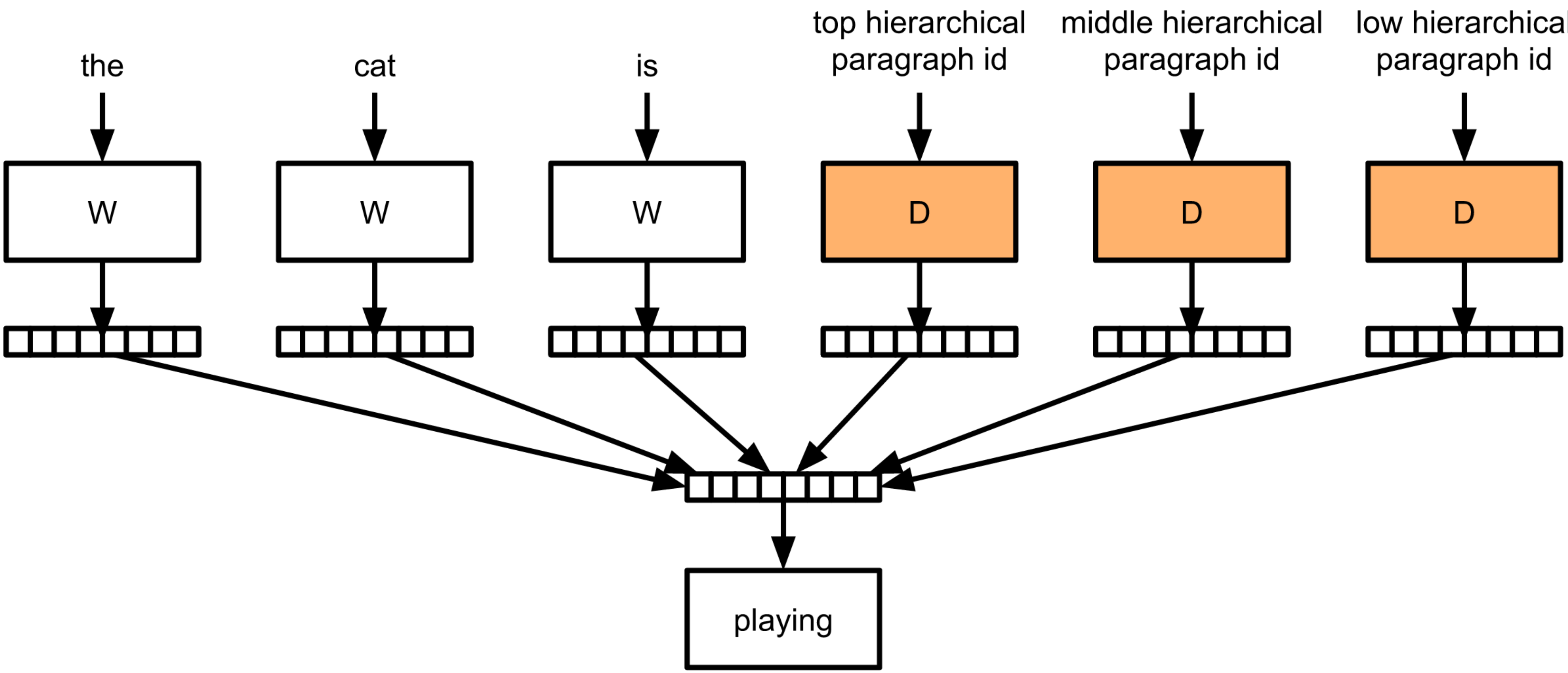
- High-level
    - Author
    - Topic
    - Tag
- Low-level
    - Chapter
    - Paragraph
    - Sentence
    - Clauses


BBC
News
Sport
Weather
Earth
Future
Shop
TV
Radio
More…
Search
Main Category
SPORT
GOLF
Subcategory
Home
Football
Formula 1
Cricket
Rugby U
Tennis
Golf
Athletics
Cycling
All Sport
Live Scores
Results
Calendar
Men's Rankings
Women's Rankings
Top Picks
The loneliness of the pro golfer
12 Jul 2015 GOLF
Brooks keeps lead as Rose stumbles
Image content (computer vision)
Featured in this story
Golf on the BBC
BBC iWonder: What makes the perfect swing?
Golf on BBC iPlayer
Iain Carter golf columns
European Tour
PGA Tour
LPGA Tour
Ladies European Tour
Scottish O
Tag
Rickie Fowler ready for Open after Gullane win
This story around the web


edit watch
Wikipedia's contents: Categories
General reference
Culture and the arts
Geography and places
Health and fitness
History and events
Mathematics and logic
Natural and physical sciences
People and self
Philosophy and thinking
Religion and belief systems
Society and social sciences
Technology and applied sciences
Categories (along with other features like cross-references, lists, and infoboxes) help you to find information, even if you don't know what exists or what it's called. The following list of categories of Wikipedia's coverage parallels our other lists by topic.
General reference (see in all page types)
edit watch
Main category
Main categories: Research and Library science
Reference works · Almanacs · Atlases · Biographical dictionaries · Dictionaries (online) · Directories (online) · Encyclopedias (online) · Glossaries · Handbooks and manuals · Lists · Medical manuals · Reference book stubs · Reference works in the public domain · Style guides · Trivia books · Web sites
Further research tools and topics · Academic disciplines · Archives · Books · Clients · Curricula · Databases (online) · Distance education · Grammar · Government agencies · Indices · Information · Knowledge · Libraries (digital) · Library cataloging and classification · News agencies · Periodic table · Prefixes · Reading · Research · Search engines · Suffixes · Topics · Universities and colleges · Writing
Quick links
The topmost category
List of all categories
About Wikipedia's categories
Category help
Random category
Category index
! 0–9 A B C D E F G H I J K L M
N O P Q R S T U V W X Y Z
Category
Culture and the arts (see in all page types)
edit watch
Main categories: Culture and Arts
Subcategory
Culture and Humanities · Classics · Critical theory · Cultural anthropology · Folklore · Food culture · Food and drink · Languages · Literature · Museology · Mythology · Philosophy · Popular culture · Science and culture · Traditions
Arts and Entertainment · Arts and crafts · Celebrity · Censorship in the arts · Festivals · Humor · Literature · Museums · Parties · Poetry
Performing arts · Circus · Dance · Film · Music · Opera · Storytelling · Theatre

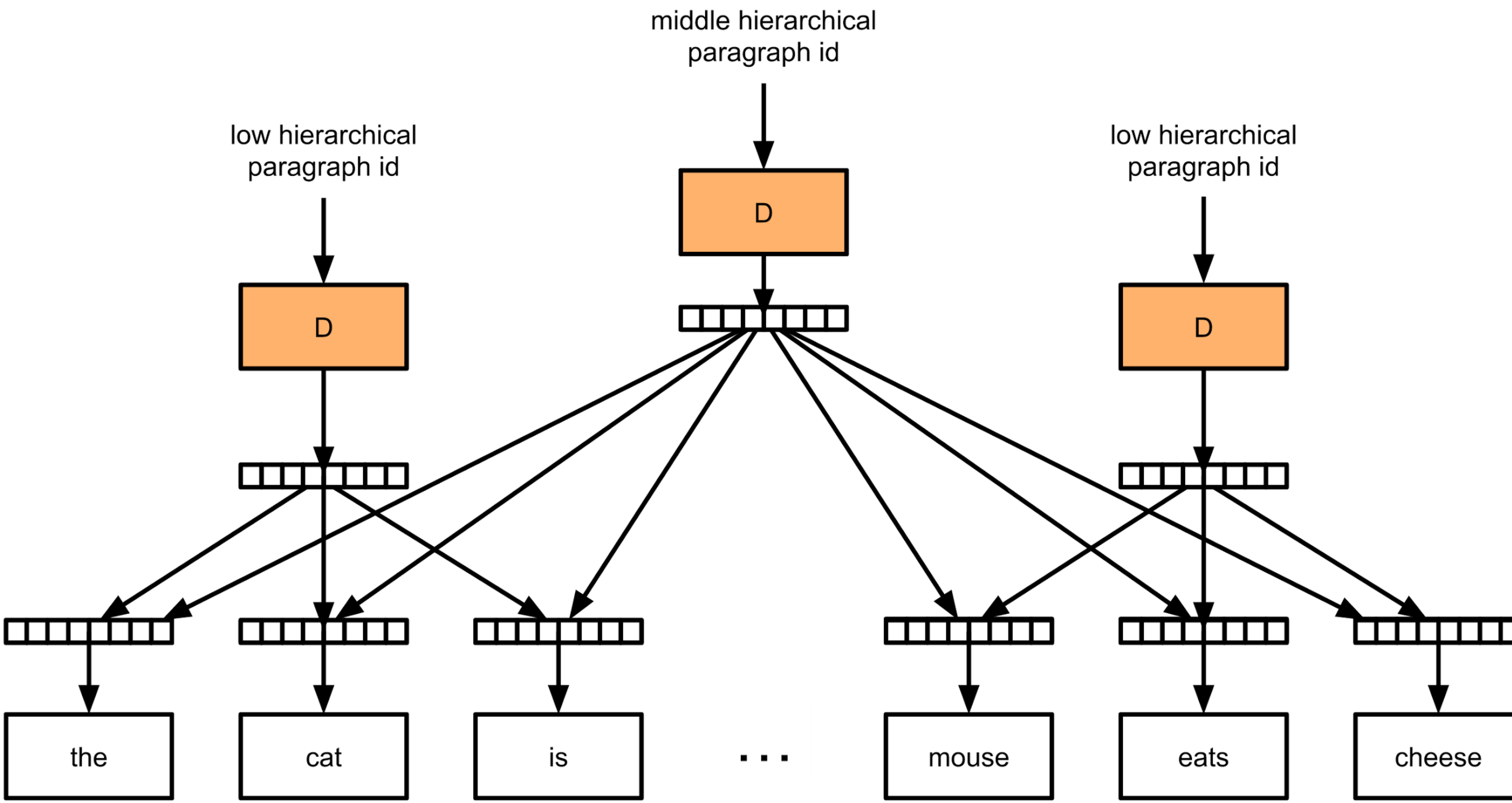

Title
Artificial intelligence
From Wikipedia, the free encyclopedia
Redirects
Chapter
History
Sentence
Paragraph
Clause
Section
Research
Goals

Goal HPV-DM



the
cat
is
top hierarchical paragraph id
middle hierarchical paragraph id
low hierarchical paragraph id
W
W
W
D
D
D
playing

Goal HPV-DBOW



Intuition

HPV = memory for larger context

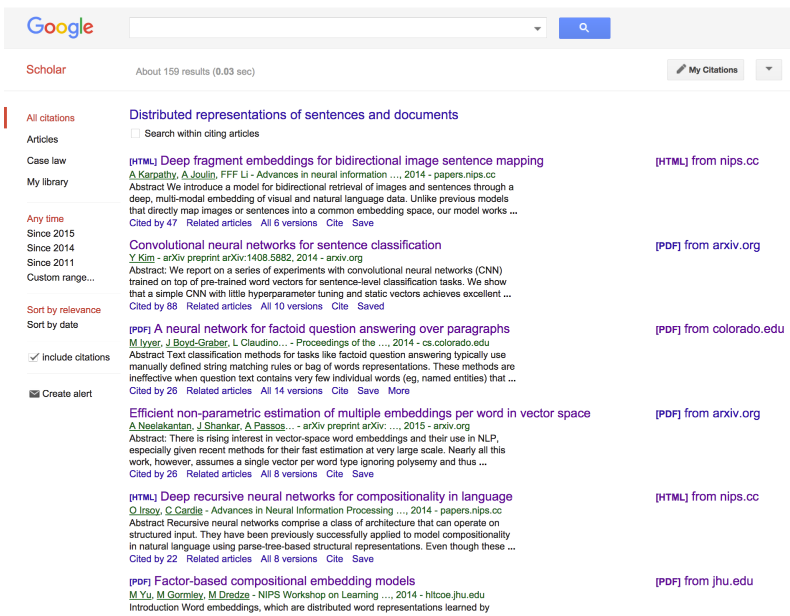
Topic #34
Chapter #2
Paragraph #20
Tag # history
Tag # neural networks

window size

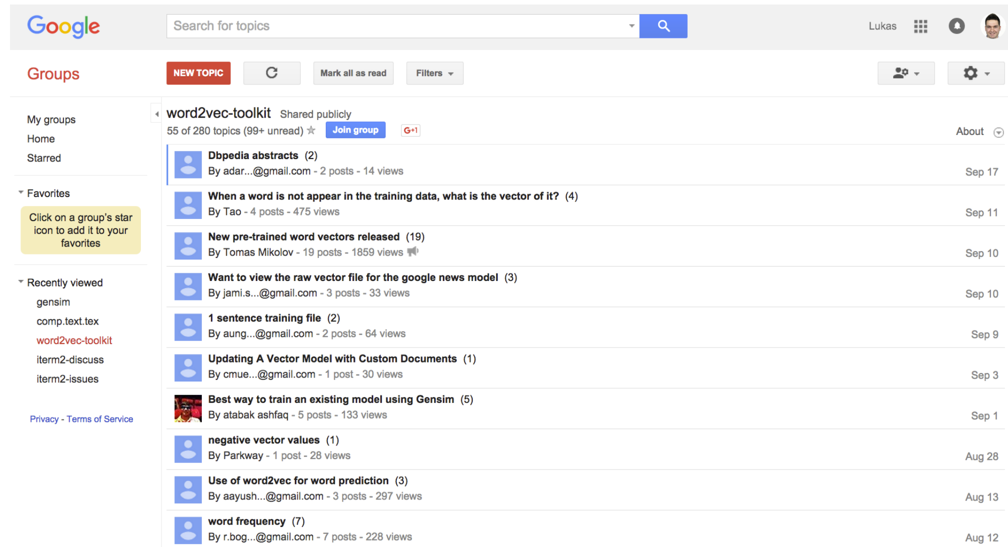
As a scientific endeavour, machine learning grew out of the quest for artificial intelligence. Already in the early days of AI as an academic discipline, some researchers were interested in having machines learn from data. They attempted to approach the problem with various symbolic methods, as well as what were then termed "neural networks"; these were mostly perceptrons and other models that were later found to be reinventions of the generalized linear models of statistics. Probabilistic reasoning was also employed, especially in automated medical diagnosis.[10]:488

3) Things Done

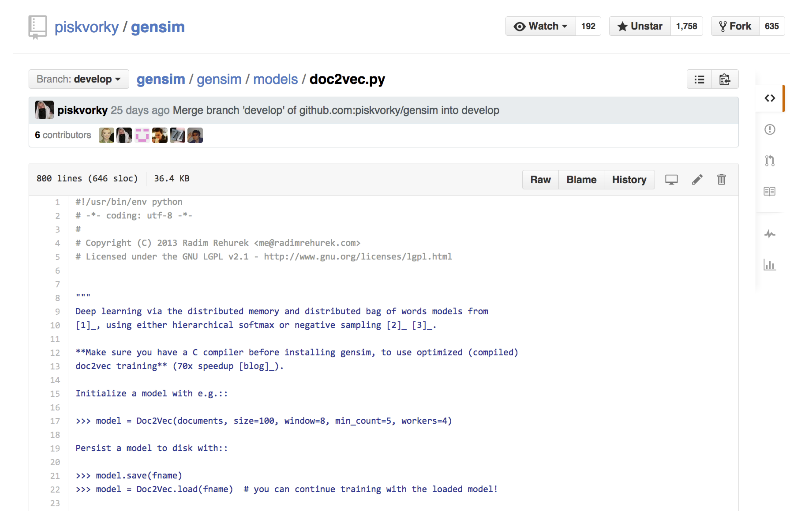
Research



Google
Scholar
About 159 results (0.03 sec)
My Citations
All citations
Articles
Case law
My library
Any time
Since 2015
Since 2014
Since 2011
Custom range...
Sort by relevance
Sort by date
include citations
Create alert
Distributed representations of sentences and documents
Search within citing articles
[HTML] Deep fragment embeddings for bidirectional image sentence mapping
A Karpathy, A Joulin, FFF Li - Advances in neural information ..., 2014 - papers.nips.cc
Abstract We introduce a model for bidirectional retrieval of images and sentences through a deep, multi-modal embedding of visual and natural language data. Unlike previous models that directly map images or sentences into a common embedding space, our model works ...
Cited by 47 Related articles All 6 versions Cite Save
[HTML] from nips.cc
Convolutional neural networks for sentence classification
Y Kim - arXiv preprint arXiv:1408.5882, 2014 - arxiv.org
Abstract: We report on a series of experiments with convolutional neural networks (CNN) trained on top of pre-trained word vectors for sentence-level classification tasks. We show that a simple CNN with little hyperparameter tuning and static vectors achieves excellent ...
Cited by 88 Related articles All 10 versions Cite Saved
[PDF] from arxiv.org
[PDF] A neural network for factoid question answering over paragraphs
M Iyyer, J Boyd-Graber, L Claudino... - Proceedings of the ..., 2014 - cs.colorado.edu
Abstract Text classification methods for tasks like factoid question answering typically use manually defined string matching rules or bag of words representations. These methods are ineffective when question text contains very few individual words (eg, named entities) that ...
Cited by 26 Related articles All 14 versions Cite Save More
[PDF] from colorado.edu
Efficient non-parametric estimation of multiple embeddings per word in vector space
A Neelakantan, J Shankar, A Passos... - arXiv preprint arXiv: ..., 2015 - arxiv.org
Abstract: There is rising interest in vector-space word embeddings and their use in NLP, especially given recent methods for their fast estimation at very large scale. Nearly all this work, however, assumes a single vector per word type ignoring polysemy and thus ...
Cited by 26 Related articles All 8 versions Cite Save
[PDF] from arxiv.org
[HTML] Deep recursive neural networks for compositionality in language
O Irsoy, C Cardie - Advances in Neural Information Processing ..., 2014 - papers.nips.cc
Abstract Recursive neural networks comprise a class of architecture that can operate on structured input. They have been previously successfully applied to model compositionality in natural language using parse-tree-based structural representations. Even though these ...
Cited by 22 Related articles All 8 versions Cite Save
[HTML] from nips.cc
[PDF] Factor-based compositional embedding models
M Yu, M Gormley, M Dredze - NIPS Workshop on Learning ..., 2014 - hltcoe.jhu.edu
Introduction Word embeddings, which are distributed word representations learned by
[PDF] from jhu.edu



Google
Search for topics
Lukas
Groups
NEW TOPIC
Mark all as read
Filters
My groups
Home
Starred
Favorites
Click on a group's star icon to add it to your favorites
Recently viewed
gensim
comp.text.tex
word2vec-toolkit
iterm2-discuss
iterm2-issues
Privacy - Terms of Service
word2vec-toolkit Shared publicly
55 of 280 topics (99+ unread)
Join group
About
Dbpedia abstracts (2)
By adar...@gmail.com - 2 posts - 14 views
Sep 17
When a word is not appear in the training data, what is the vector of it? (4)
By Tao - 4 posts - 475 views
Sep 11
New pre-trained word vectors released (19)
By Tomas Mikolov - 19 posts - 1859 views
Sep 10
Want to view the raw vector file for the google news model (3)
By jami.s...@gmail.com - 3 posts - 33 views
Sep 10
1 sentence training file (2)
By aung...@gmail.com - 2 posts - 64 views
Sep 9
Updating A Vector Model with Custom Documents (1)
By cmue...@gmail.com - 1 post - 30 views
Sep 3
Best way to train an existing model using Gensim (5)
By atabak ashfaq - 5 posts - 133 views
Sep 1
negative vector values (1)
By Parkway - 1 post - 28 views
Aug 28
Use of word2vec for word prediction (3)
By aayush...@gmail.com - 3 posts - 297 views
Aug 13
word frequency (7)
By r.bog...@gmail.com - 7 posts - 228 views
Aug 12



piskvorky / gensim
Watch 192
Unstar 1,758
Fork 635
Branch: develop
gensim / gensim / models / doc2vec.py
piskvorky 25 days ago Merge branch 'develop' of github.com:piskvorky/gensim into develop
6 contributors
800 lines (646 sloc) 36.4 KB
Raw
Blame
History
#!/usr/bin/env python
-*- coding: utf-8 -*-
#
Copyright (C) 2013 Radim Rehurek <me@radimrehurek.com>
Licensed under the GNU LGPL v2.1 - http://www.gnu.org/licenses/lgpl.html
"""
Deep learning via the distributed memory and distributed bag of words models from
[1]_, using either hierarchical softmax or negative sampling [2]_ [3]_.
**Make sure you have a C compiler before installing gensim, to use optimized (compiled)
doc2vec training** (70x speedup [blog]_).
Initialize a model with e.g.::
>>> model = Doc2Vec(documents, size=100, window=8, min_count=5, workers=4)
Persist a model to disk with::
>>> model.save(fname)
>>> model = Doc2Vec.load(fname) # you can continue training with the loaded model!
The model can also be instantiated from an existing file on disk in the word2vec C format::

HPV Implementations

Abbreviation
NO-HPV
HPV-TOP
HPV-TOP-PAR
HPV-TOP-PAR-SENT
HPV-PAR
HPV-PAR-SENT
HPV-PAR-SENT-SUB
HPV-PAR-SENT-SUBNV

Hyperparameter Search

Name	*Range / Options*	*Best / Chosen*
Epochs	$[5,\ 50]$	20
HPV Hierarchies	TOP, PAR, SENT, SUB, SUBNV	TOP
Word Vector Dimensionality	$[16,\ 2000]$	200, 48
Window Size	$[5,\ 25]$	10
Negative Sampling	$[5,\ 30]$	25
Frequent Word Downsampling HPV-DM	$[10^{-10},\ 0.1]$	10^{-5}
Frequent Word Downsampling HPV-DBOW	$[10^{-10},\ 0.1]$	10^{-3}
Learning Rate Type	exp, lin	exp
Classifier	RBF, LOG1/2, SVC	SVC

Experiments Runner

Summary

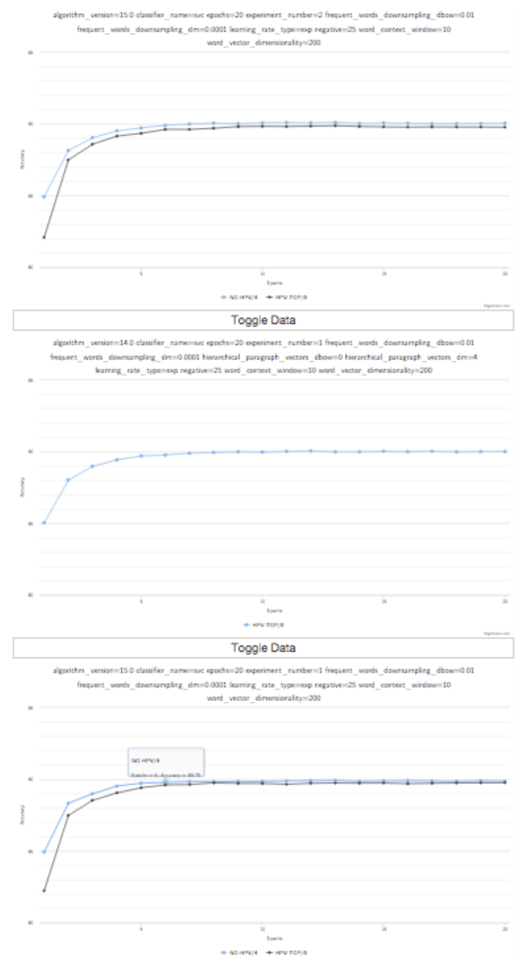
Pending: 0
Ended: 21858
Failed: 1740

10 days ago	89.372	{"negative":25, "epochs_total":20, "algorithm_version":15.0, "experiment_number":2, "learning_rate_type":"exp", "word_context_window":10, "word_vector_dimensionality":48, "frequent_words_downsampling_dm":0.0001, "hierarchical_paragraph_vectors":"4/0", "frequent_words_downsampling_dbow":0.01, "epochs":16, "classifier_c":0.0195, "tfid_features":0, "classifier_name":"svc", "classifier_penalty":"l2"}	Lukass-MacBook-Pro-2.local	Show	Restart
10 days ago	89.32	{"negative":25, "epochs_total":20, "algorithm_version":15.0, "experiment_number":2, "learning_rate_type":"exp", "word_context_window":10, "word_vector_dimensionality":48, "frequent_words_downsampling_dm":0.0001, "hierarchical_paragraph_vectors":"4/0", "frequent_words_downsampling_dbow":0.01, "epochs":12, "classifier_c":0.0195, "tfid_features":0, "classifier_name":"svc", "classifier_penalty":"l2"}	Lukass-MacBook-Pro-2.local	Show	Restart
10 days ago	89.364	{"negative":25, "epochs_total":20, "algorithm_version":15.0, "experiment_number":2, "learning_rate_type":"exp", "word_context_window":10, "word_vector_dimensionality":48, "frequent_words_downsampling_dm":0.0001, "hierarchical_paragraph_vectors":"4/0", "frequent_words_downsampling_dbow":0.01, "epochs":15, "classifier_c":0.0195, "tfid_features":0, "classifier_name":"svc", "classifier_penalty":"l2"}	Lukass-MacBook-Pro-2.local	Show	Restart
10 days ago	89.22	{"negative":25, "epochs_total":20, "algorithm_version":15.0, "experiment_number":2, "learning_rate_type":"exp", "word_context_window":10, "word_vector_dimensionality":48, "frequent_words_downsampling_dm":0.0001, "hierarchical_paragraph_vectors":"4/0", "frequent_words_downsampling_dbow":0.01, "epochs":10, "classifier_c":0.0195, "tfid_features":0, "classifier_name":"svc", "classifier_penalty":"l2"}	Lukass-MacBook-Pro-2.local	Show	Restart

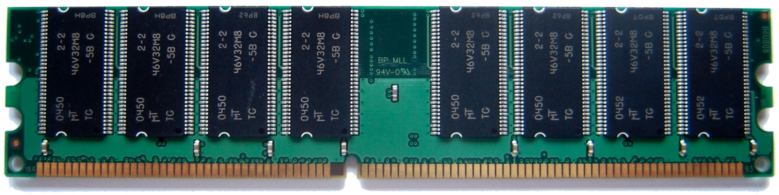
Simple Job Runner

Epochs vs Score Grouped by Hierarchical paragraph vectors

algorithm_version classifier_c classifier_name classifier_penalty epochs epochs_total experiment_number force_restart frequent_words_downsampling frequent_words_downsampling_dbow frequent_words_downsampling_dm hierarchical_paragraph_vectors hierarchical_paragraph_vectors_dbow hierarchical_paragraph_vectors_dm invalid learning_rate_type negative no_host_name tfid_features use_converted word_context_window word_vector_dimensionality



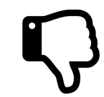
Efficiency



4) One Real World Problem

Reviews

Review 1: “The script for this movie was probably found in a hair-ball recently coughed up by a really old dog. Mostly an **amateur film** with **lame** FX. For you Zeta-Jones fanatics: she has the credibility of one Mr. Binks.”



Review 2: “This movie could have been **very good**, but comes up way short. Cheesy special effects and so-so acting. I could have looked past that if the story wasn't so lousy. If there was more of a background story, it would have been better. The plot centers around an evil Druid witch who is linked to this woman who gets migraines. The movie **drags** on and on and never clearly explains anything, it just keeps plodding on. Christopher Walken has a part, but it is completely senseless, as is most of the movie. This movie had potential, but it looks like some really **bad** made for TV movie. I would **avoid** this movie.”

Review 3: “What happens when an army of wetbacks, towelheads, and Godless Eastern European commies gather their forces south of the border? Gary Busey kicks their butts, of course. Another **laughable** example of Reagan-era cultural fallout, **Bulletproof** wastes a decent supporting cast headed by L Q Jones and Thalmus Rasulala.”

Reviews

Review 1: “The script for this movie was probably found in a hair-ball recently coughed up by a really old dog. Mostly an **amateur film** with **lame** FX. For you Zeta-Jones fanatics: she has the credibility of one Mr. Binks.”

Review 2: “This movie could have been **very good**, but comes up way short. Cheesy special effects and so-so acting. I could have looked past that if the story wasn't so lousy. If there was more of a background story, it would have been better. The plot centers around an evil Druid witch who is linked to this woman who gets migraines. The movie **drags** on and on and never clearly explains anything, it just keeps plodding on. Christopher Walken has a part, but it is completely senseless, as is most of the movie. This movie had potential, but it looks like some really **bad** made for TV movie. I would **avoid** this movie.”

Review 3: “What happens when an army of wetbacks, towelheads, and Godless Eastern European commies gather their forces south of the border? Gary Busey kicks their butts, of course. Another **laughable** example of Reagan-era cultural fallout, **Bulletproof** wastes a decent supporting cast headed by L Q Jones and Thalmus Rasulala.”

Review 4: “Although I generally do not like remakes believing that remakes are waste of time; this film is an exception. I didn't actually know so far until reading the previous comment that this wa s a remake, so my opinion is purely about the actual film and not a comparison.

The story and the way it is written is no question: it is Capote. There is no need for more words.

The play of Anthony Edwards and Eric Roberts is **superb**. I have seen some movies with them, each in one or the other. I was certain that they are good actors and in case of Eric I always wondered why his sister is the number 1 famous star and not her brother. This time this certainty is raised to fact, no question. His play, just as well as the play of Mr. Edwards is clearly the top of all their profession.

I recommend th is film to be on your **top 50 films** to see and keep on your DVD shelves.”

Reviews

Review 1: “The script for this movie was probably found in a hair-ball recently coughed up by a really old dog. Mostly an **amateur film** with **lame** FX. For you Zeta-Jones fanatics: she has the credibility of one Mr. Binks.”

Review 2: “This movie could have been **very good**, but comes up way short. Cheesy special effects and so-so acting. I could have looked past that if the story wasn't so lousy. If there was more of a background story, it would have been better. The plot centers around an evil Druid witch who is linked to this woman who gets migraines. The movie **drags** on and on and never clearly explains anything, it just keeps plodding on. Christopher Walken has a part, but it is completely senseless, as is most of the movie. This movie had potential, but it looks like some really **bad** made for TV movie. I would **avoid** this movie.”

Review 3: “What happens when an army of wetbacks, towelheads, and Godless Eastern European commies gather their forces south of the border? Gary Busey kicks their butts, of course. Another **laughable** example of Reagan-era cultural fallout, **Bulletproof** wastes a decent supporting cast headed by L Q Jones and Thalmus Rasulala.”

Review 4: “Although I generally do not like remakes believing that remakes are waste of time; this film is an exception. I didn't actually know so far until reading the previous comment that this wa s a remake, so my opinion is purely about the actual film and not a comparison.

The story and the way it is written is no question: it is Capote. There is no need for more words.

The play of Anthony Edwards and Eric Roberts is **superb**. I have seen some movies with them, each in one or the other. I was certain that they are good actors and in case of Eric I always wondered why his sister is the number 1 famous star and not her brother. This time this certainty is raised to fact, no question. His play, just as well as the play of Mr. Edwards is clearly the top of all their profession.

I recommend th is film to be on your **top 50 films** to see and keep on your DVD shelves.”

Task

Given: movie reviews

- pos: reviewer likes the movie
- neg: reviewer does not like the movie

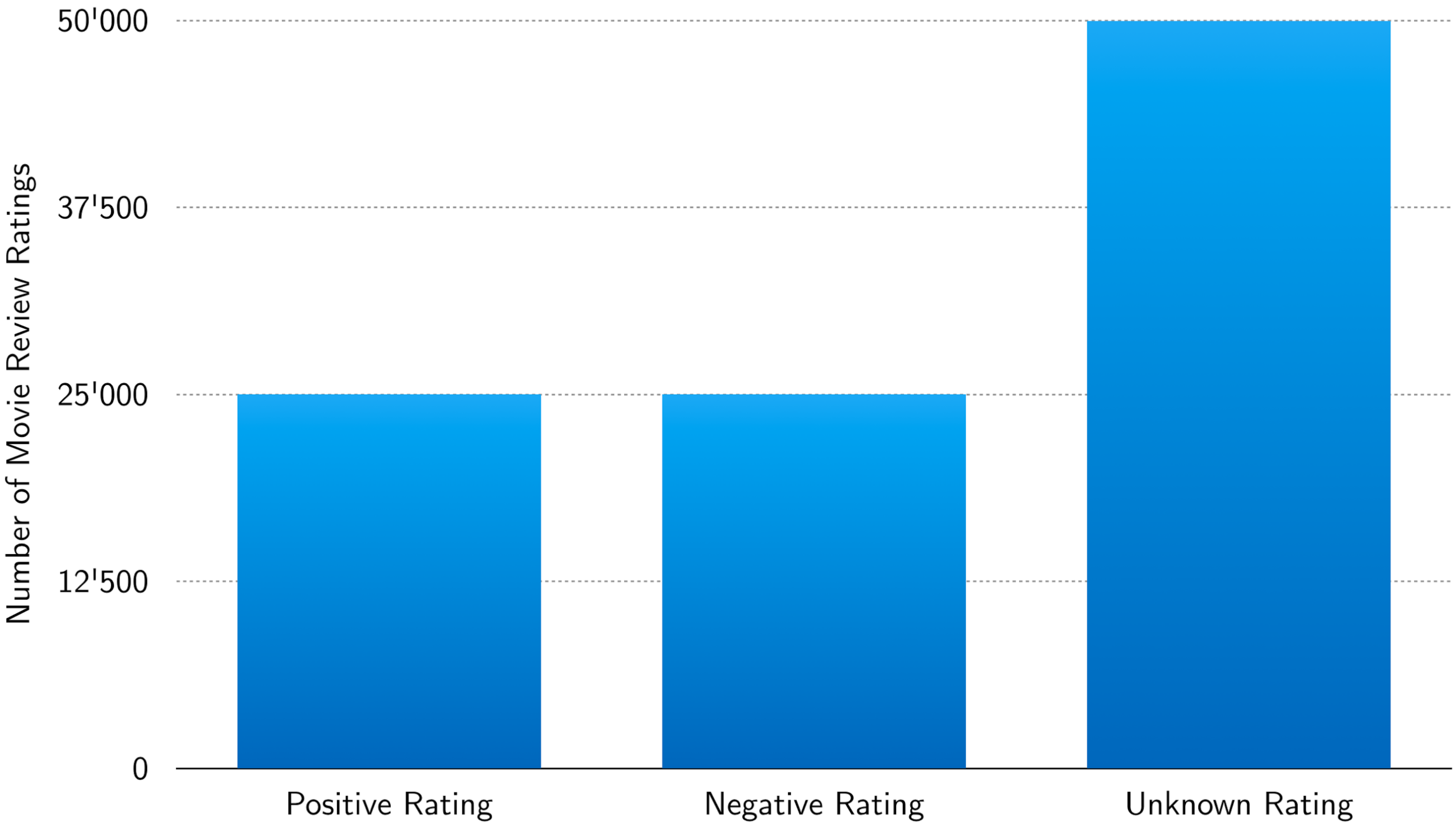
Task: Sentiment analysis

Predict sentiment from text

- optimize mean accuracy

$$\text{accuracy} = \frac{1}{\dim(R)} \sum_{r \in R} \begin{cases} 1 & \text{when } r_{\text{predicted sentiment}} = r_{\text{actual sentiment}} \\ 0 & \text{otherwise} \end{cases}$$

Data Distribution



Split

Learn embeddings for 100’000 reviews

- 1 review vector per movie review

Split labeled reviews

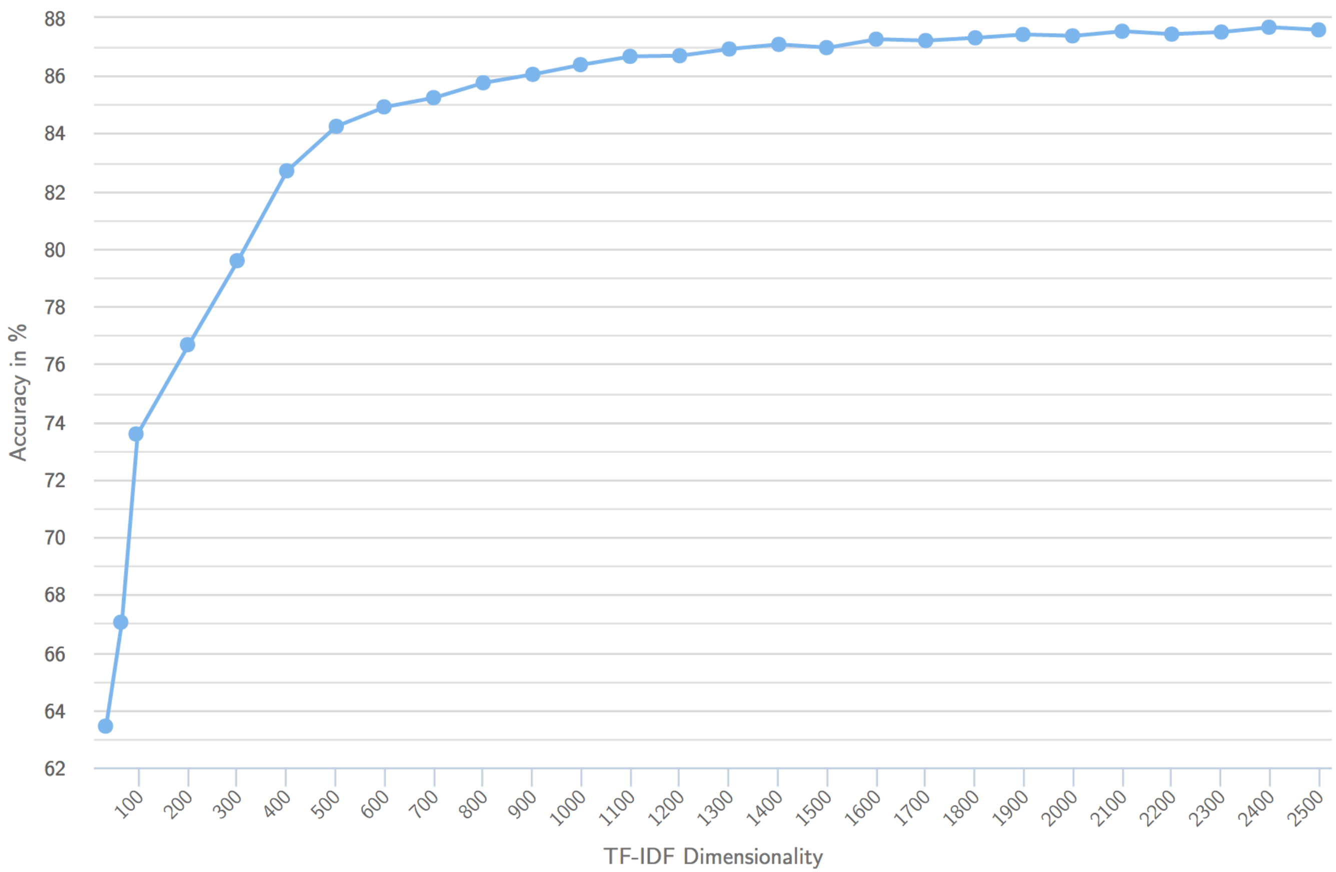
- 25’000 for training
- 25’000 for testing

Algorithm

1) preprocessing
2) for epoch in epochs
    a) train review vectors
        i) HPV-DM and HPV-DBOW
        ii) concatenate
    b) SVC with linear kernel
        i) learn sentiment
        ii) predict sentiment
    c) accuracy

5) Results

Results - TF-IDF Baseline



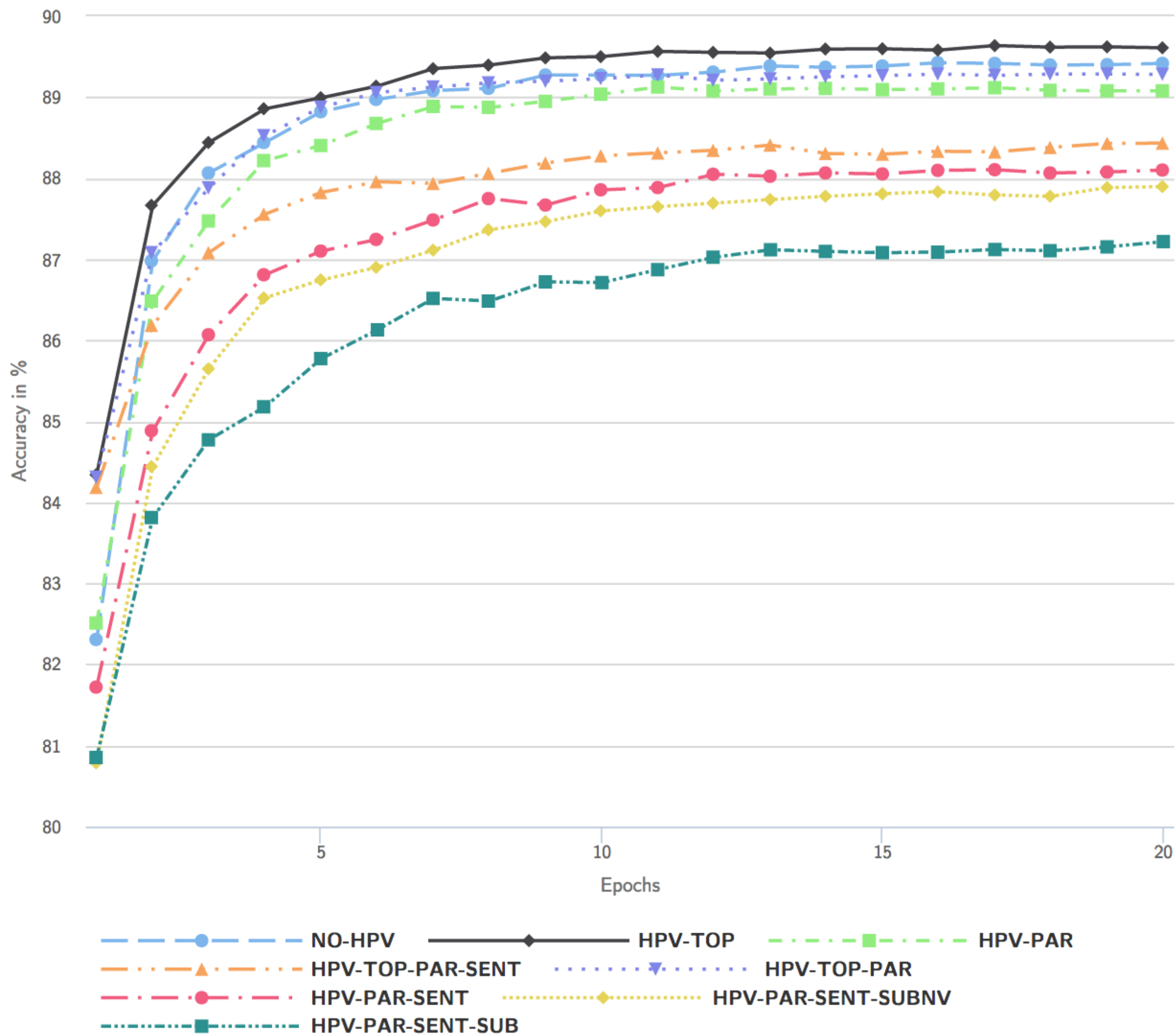
Results - HPV Implementations

Abbreviation	*Hierarchy*
TOP	Topics
PAR	Paragraphs
SENT	Sentences
SUB	Sub-sentences
SUBNV	Sub-sentences (but without training the sub-sentence vectors)

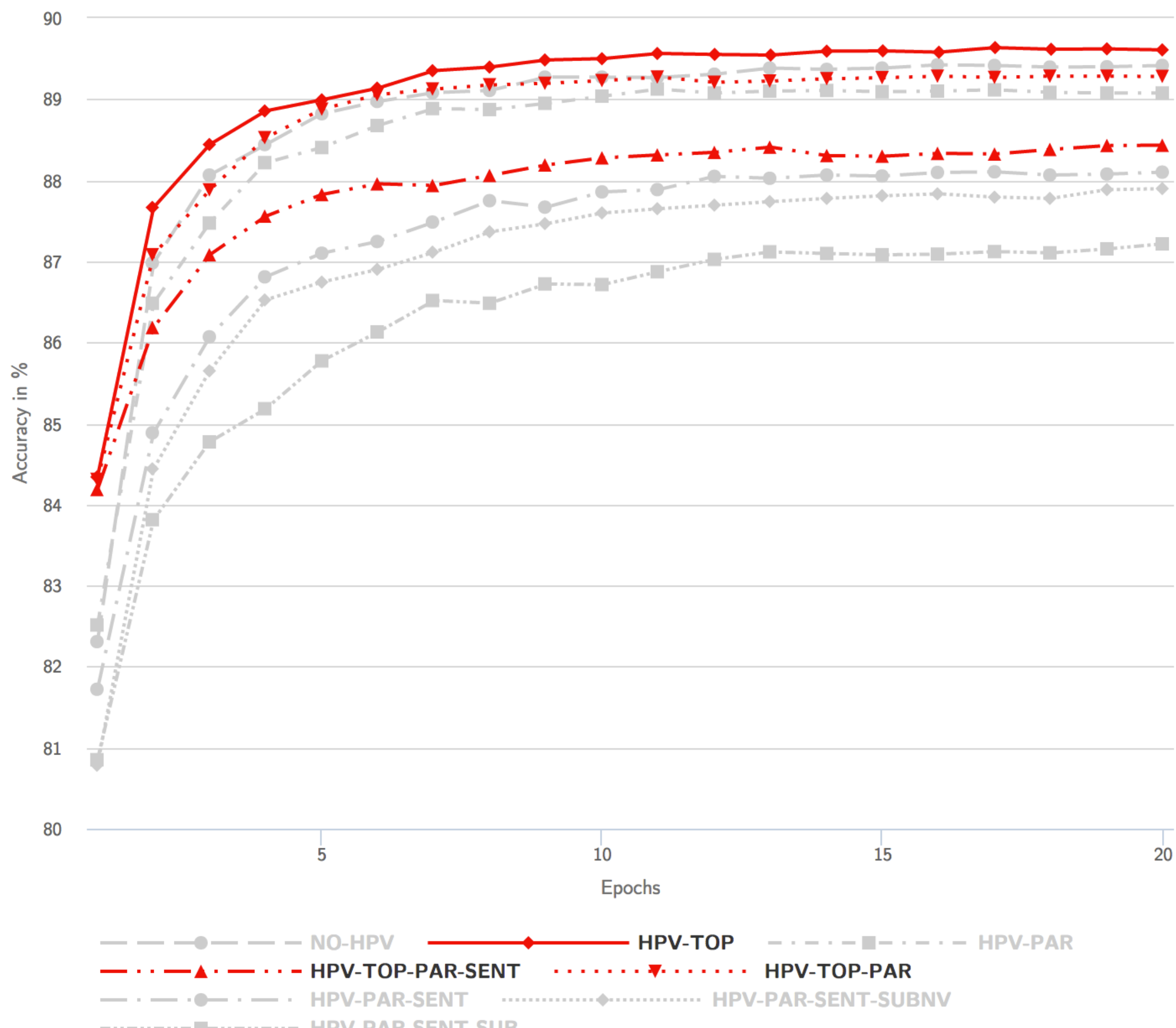
Results - HPV Implementations

Abbreviation	*Hierarchies*
NO-HPV	None
HPV-TOP	Topics
HPV-TOP-PAR	Topics, Paragraphs
HPV-TOP-PAR-SENT	Topics, Paragraphs, Sentences
HPV-PAR	Paragraphs
HPV-PAR-SENT	Paragraphs, Sentences
HPV-PAR-SENT-SUB	Paragraphs, Sentences, Sub-sentences
HPV-PAR-SENT-SUBNV	Paragraphs, Sentences, Sub-sentences (but without training the sub-sentence vectors)

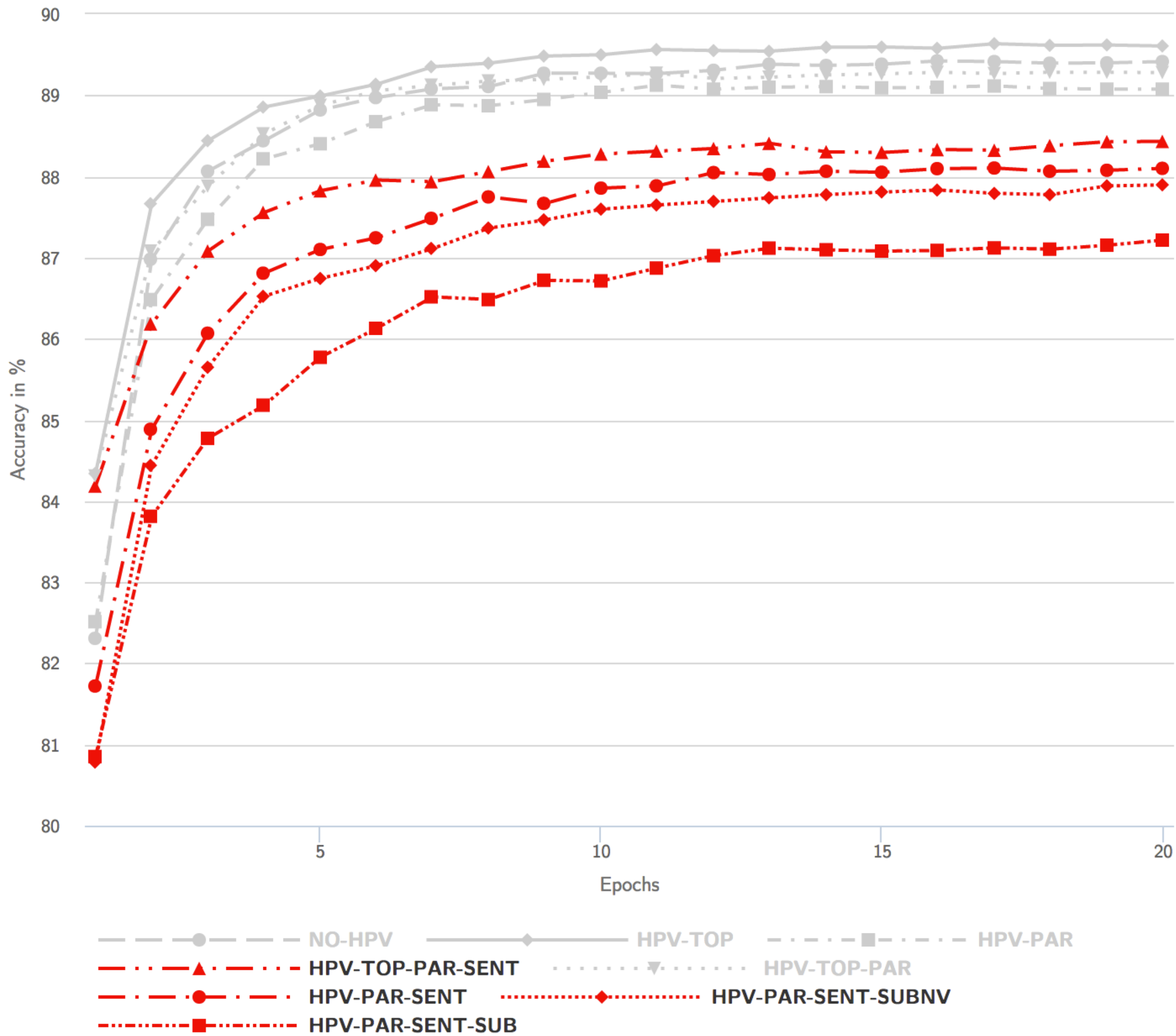
Results - HPV Implementations Using 96 Dimensions



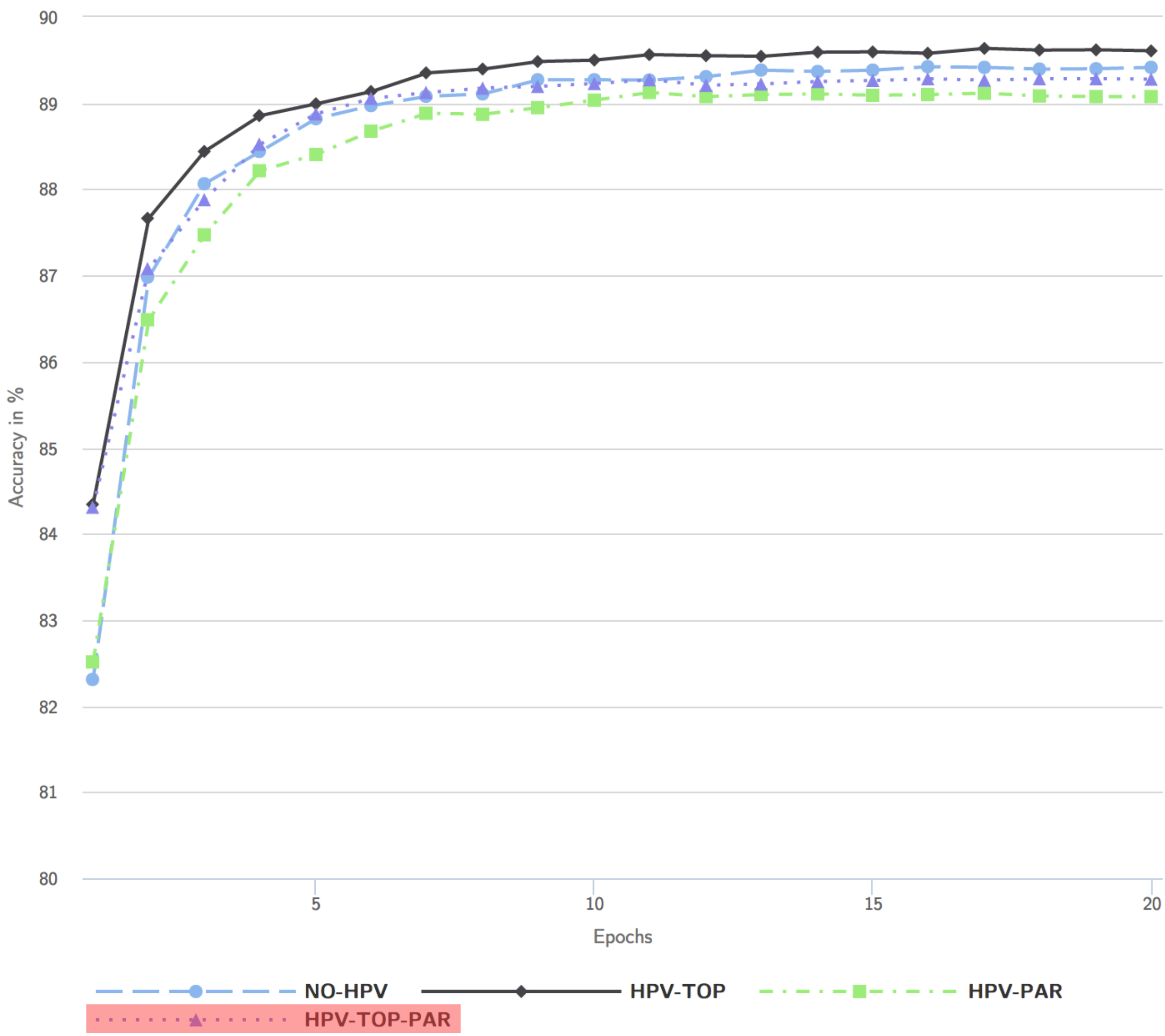
Results - HPV Implementations Using 96 Dimensions



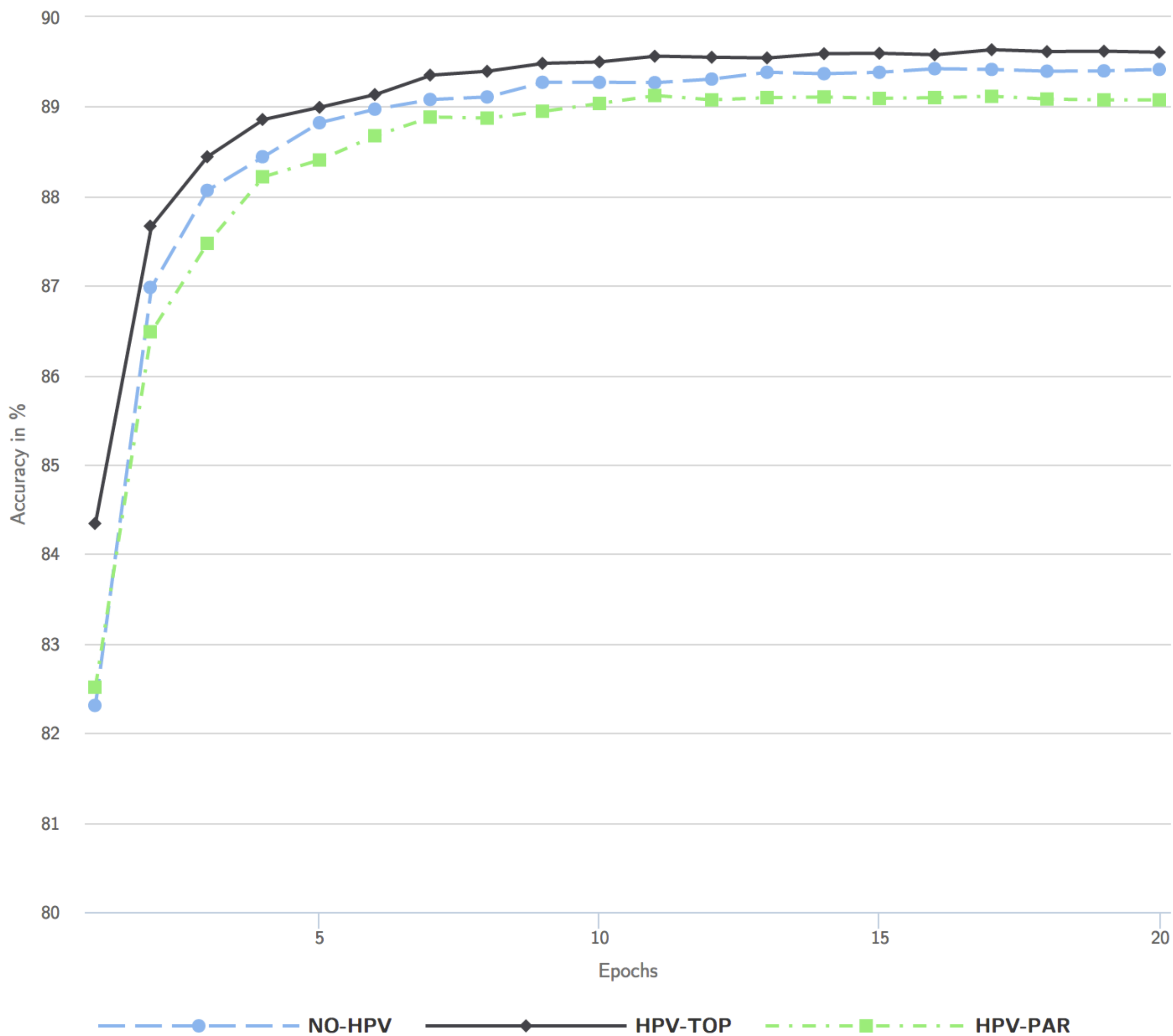
Results - HPV Implementations Using 96 Dimensions



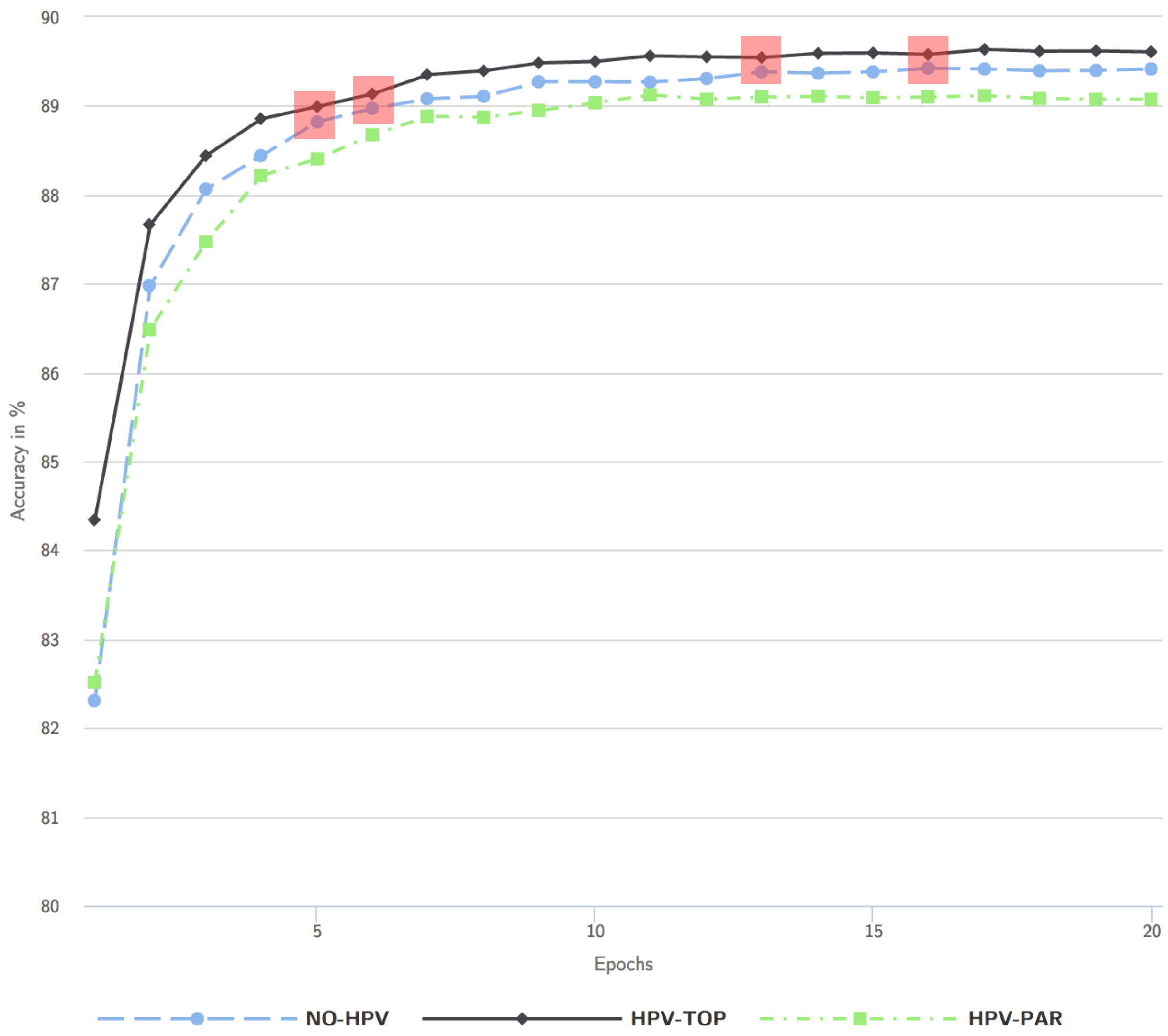
Results - HPV Implementations Using 96 Dimensions



Results - HPV Implementations Using 96 Dimensions



Results - HPV Implementations Using 96 Dimensions



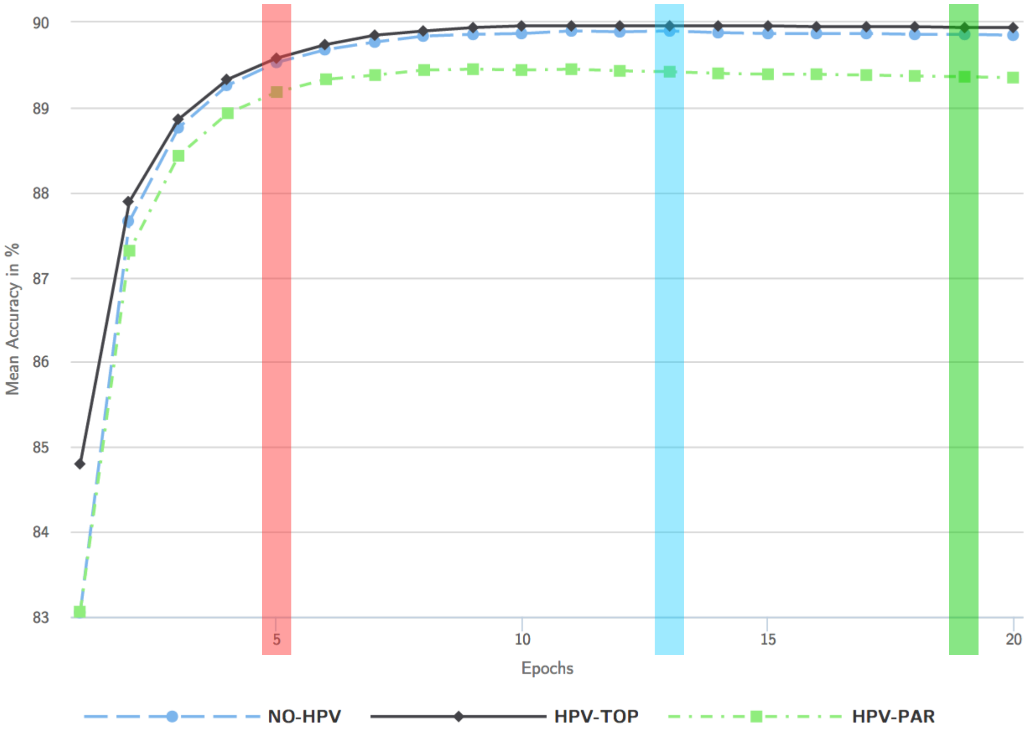
Results - 400 Dimensions

Significance tests, 30 experiments

Results - 400 Dimensions

Significance tests, 30 experiments

	Accuracy [%]		
Epoch	*HPV-NO*	*HPV-TOP*	*HPV-PAR*
1	83.05	*84.8	83.06
2	87.66	*87.89	87.31
3	88.76	*88.86	88.43
4	89.26	*89.33	88.93
5	89.53	89.58	89.18
6	89.68	*89.74	89.33
7	89.77	*89.85	89.38
8	89.84	*89.9	89.44
9	89.86	*89.94	89.45
10	89.87	*89.96	89.44
11	89.9	*89.96	89.45
12	89.89	*89.96	89.43
13	89.9	89.96	89.42
14	89.88	*89.96	89.4
15	89.87	*89.96	89.39
16	89.87	*89.95	89.39
17	89.87	*89.95	89.38
18	89.86	*89.95	89.37
19	89.86	89.94	89.36
20	89.85	*89.94	89.35



Results - Efficiency

Results - Runtime

HPV	*Model*	*Average Duration [s]*	*Standard Deviation*
NO-HPV	HPV-DBOW	23.71	0.36
	HPV-DM	18.01	0.65
HPV-TOP	HPV-DBOW	65.61	0.61
	HPV-DM	18.7	0.42
HPV-TOP-PAR	HPV-DBOW	88.85	0.72
	HPV-DM	27.2	0.55
HPV-TOP-PAR-SENT	HPV-DBOW	118.62	9.19
	HPV-DM	89.98	4.3
HPV-PAR	HPV-DBOW	45.91	0.83
	HPV-DM	26.98	0.23
HPV-PAR-SENT	HPV-DBOW	76.63	3.56
	HPV-DM	78.85	0.22
HPV-PAR-SENT-SUBNV	HPV-DBOW	111.79	5.7
	HPV-DM	135.13	1.17
HPV-PAR-SENT-SUB	HPV-DBOW	120.08	2.84
	HPV-DM	140.72	1.6

Results - Runtime

HPV	*Model*	*Average Duration [s]*	*Standard Deviation*
NO-HPV	HPV-DBOW	23.71	0.36
	HPV-DM	18.01	0.65
HPV-TOP	HPV-DBOW	65.61	0.61
	HPV-DM	18.7	0.42
HPV-TOP-PAR	HPV-DBOW	88.85	0.72
	HPV-DM	27.2	0.55
HPV-TOP-PAR-SENT	HPV-DBOW	118.62	9.19
	HPV-DM	89.98	4.3
HPV-PAR	HPV-DBOW	45.91	0.83
	HPV-DM	26.98	0.23
HPV-PAR-SENT	HPV-DBOW	76.63	3.56
	HPV-DM	78.85	0.22
HPV-PAR-SENT-SUBNV	HPV-DBOW	111.79	5.7
	HPV-DM	135.13	1.17
HPV-PAR-SENT-SUB	HPV-DBOW	120.08	2.84
	HPV-DM	140.72	1.6

Results - Runtime

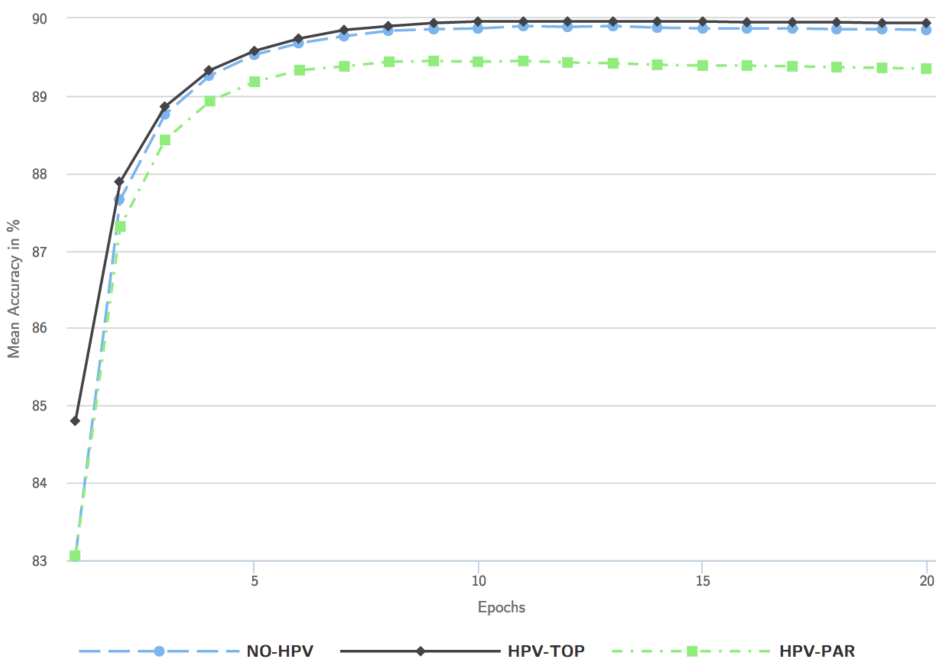
HPV	*Model*	*Average Duration [s]*	*Standard Deviation*
NO-HPV	HPV-DBOW	23.71	0.36
	HPV-DM	18.01	0.65
HPV-TOP	HPV-DBOW	65.61	0.61
	HPV-DM	18.7	0.42
HPV-TOP-PAR	HPV-DBOW	88.85	0.72
	HPV-DM	27.2	0.55
HPV-TOP-PAR-SENT	HPV-DBOW	118.62	9.19
	HPV-DM	89.98	4.3
HPV-PAR	HPV-DBOW	45.91	0.83
	HPV-DM	26.98	0.23
HPV-PAR-SENT	HPV-DBOW	76.63	3.56
	HPV-DM	78.85	0.22
HPV-PAR-SENT-SUBNV	HPV-DBOW	111.79	5.7
	HPV-DM	135.13	1.17
HPV-PAR-SENT-SUB	HPV-DBOW	120.08	2.84
	HPV-DM	140.72	1.6

Results - Runtime

HPV	Total Average Duration [s]	[%]
NO-HPV	41.72	100
HPV-TOP	84.31	202
HPV-TOP-PAR	116.05	278
HPV-TOP-PAR-SENT	208.60	500
HPV-PAR	72.89	175
HPV-PAR-SENT	155.48	373
HPV-PAR-SENT-SUBNV	246.92	592
HPV-PAR-SENT-SUB	260.80	625

Results - Runtime

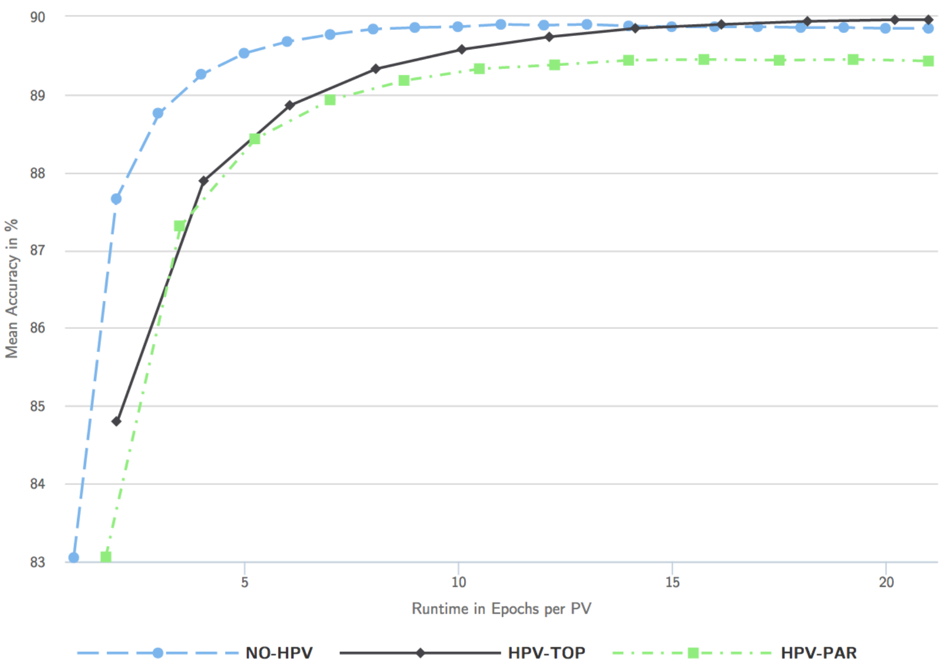
HPV	*Total Average Duration* *[s]*	*[%]*
NO-HPV	41.72	100
HPV-TOP	84.31	202
HPV-TOP-PAR	116.05	278
HPV-TOP-PAR-SENT	208.60	500
HPV-PAR	72.89	175
HPV-PAR-SENT	155.48	373
HPV-PAR-SENT-SUBNV	246.92	592
HPV-PAR-SENT-SUB	260.80	625



Results - Runtime

Stretched according to runtime overhead

HPV	*Total Average Duration* [s]	[%]
NO-HPV	41.72	100
HPV-TOP	84.31	202
HPV-TOP-PAR	116.05	278
HPV-TOP-PAR-SENT	208.60	500
HPV-PAR	72.89	175
HPV-PAR-SENT	155.48	373
HPV-PAR-SENT-SUBNV	246.92	592
HPV-PAR-SENT-SUB	260.80	625



Results - Problems

- How did Le get 92.6%?!
  - Mikolov: “[...] I am starting to think that [Le's] results are actually not reproducible.”
  - Later, Le:
    “A simple modification to the word2vec training command should give better results. Try changing line 55 of your go.sh to this command:
    time ./word2vec -train ../alldata-id.txt -output vectors.txt -cbow 0 -size 100 -window 10 -negative 5 -hs 1 -sample 1e-3 -threads 40 -binary 0 -iter 20 -min-count 1 -sentence-vectors 1
    On my machine, I got 92.6% on the dataset. Combining with rnnlm features should get better than 93%. If you combine PV-DM and PV-DBOW, you should get similar results as well.”
  - Conclusion: unable to reproduce
    - NN classifier? Implementation (my code? gensim?)? Preprocessing? Hyperparameters? Metric? Model combinations?

[1] https://groups.google.com/forum/#!msg/word2vec-toolkit/Q49FIrNOQRo/J6KG8mUj45sJ

6) Conclusion and Future Work

Conclusion

- Novel method to improve word embeddings by using hierarchical data
- 7 different HPV implementations
  - High-level hierarchies improve word embeddings, greater execution overhead
  - Low-level hierarchies did not improve embeddings
- Initial boost when using HPV-TOP
  - Scaled with runtime overhead: no initial boost
- Accuracy improvement for sentiment analysis of IMDB movie reviews

Future Work

- Datasets / applications
- Optimize implementation
- Hyperparameters / combinations
- Model combinations
  - E.g. HPV-DM + PV-DBOW
    - → different models have different runtime
  - E.g. HPV-DM 48 dimensions + HPV-DM 200 dimensions + HPV-DM 400 dimensions
- Independent model evaluation
  - Currently: HPV-DM + HPV-DBOW
- Publish results online

Publish Code and Website?

https://github.com/lukaselmer/hierarchical-paragraph-vectors/settings

Published Code and Website

https://hpv.renuo.ch

Thank you!

Demo Simple Job Runner

https://simple-job-runner-master.herokuapp.com/

Questions / Discussion

https://hpv.renuo.ch

lukas.elmer@gmail.com