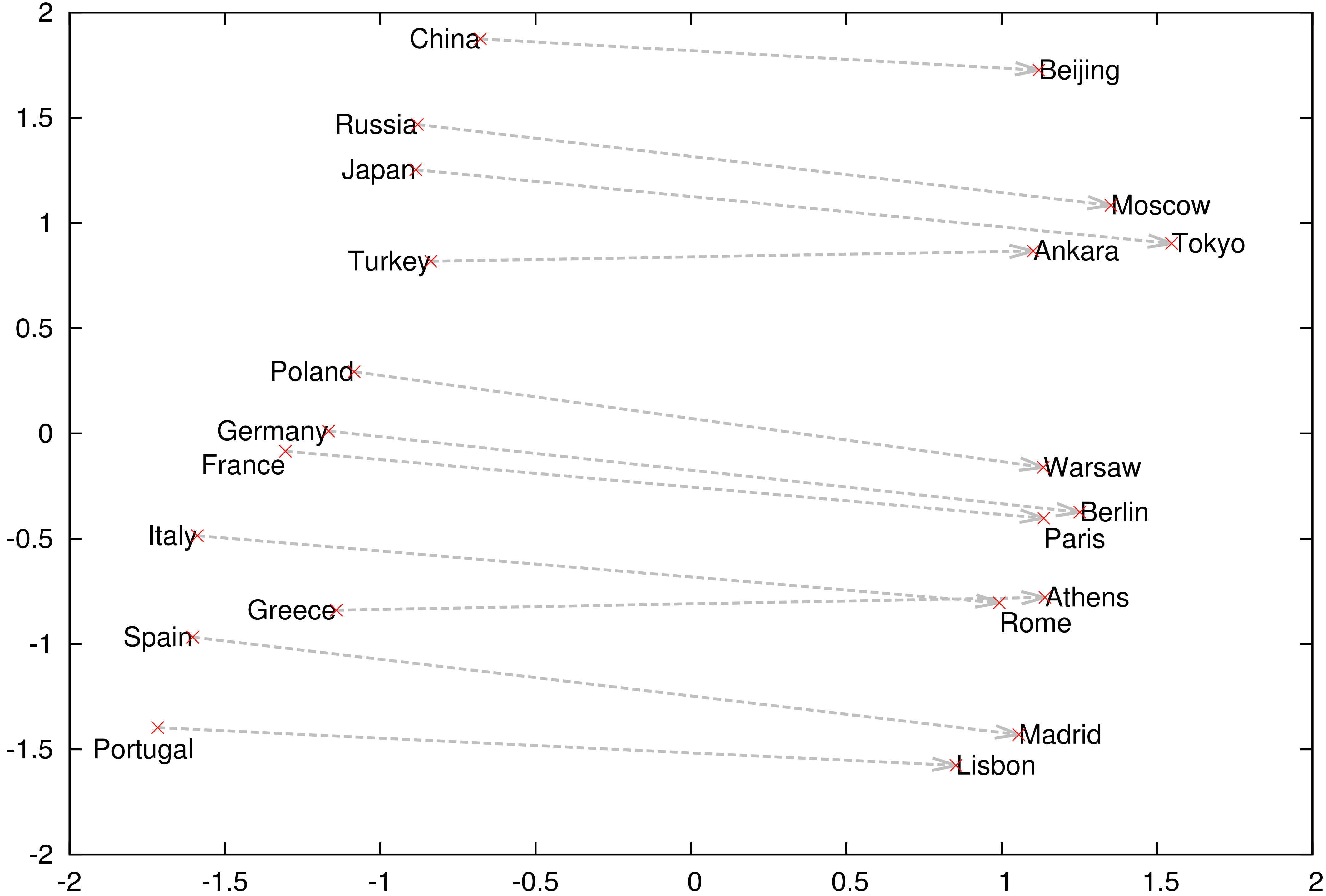
Modelling word analogies with language models

Steven Schockaert

School of Computer Science & Informatics
Cardiff University, Cardiff, UK
SchockaertS1@cardiff.ac.uk
http://users.cs.cf.ac.uk/S.Schockaert

Word vectors and analogies



T. Mikolov. Distributed representations of words and phrases and their compositionality. NIPS 2013

Why do vector differences model analogies?

Typical context words of “France”: *arrondissement, renaissance, …*

Typical context words of “capital”: *embassy, palace, …*

Typical context words of “Paris”: *arrondissement, embassy, palace, …*

Carl Allen, Timothy M. Hospedales: Analogies Explained: Towards Understanding Word Embeddings. ICML 2019: 223-231

Why do vector differences model analogies?

Typical context words of France: *arrondissement, renaissance, …*

+

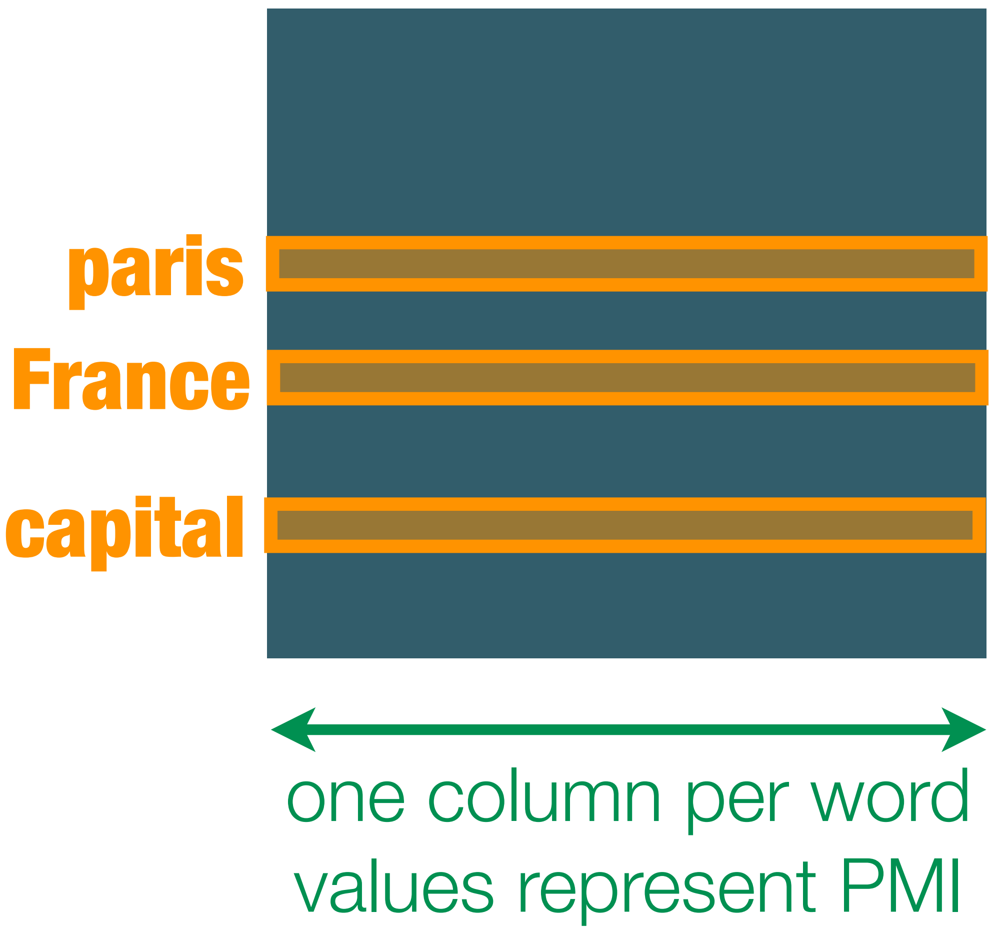
Typical context words of capital cities: *embassy, palace, …*

=

Typical context words of Paris: *arrondissement, embassy, palace, …*

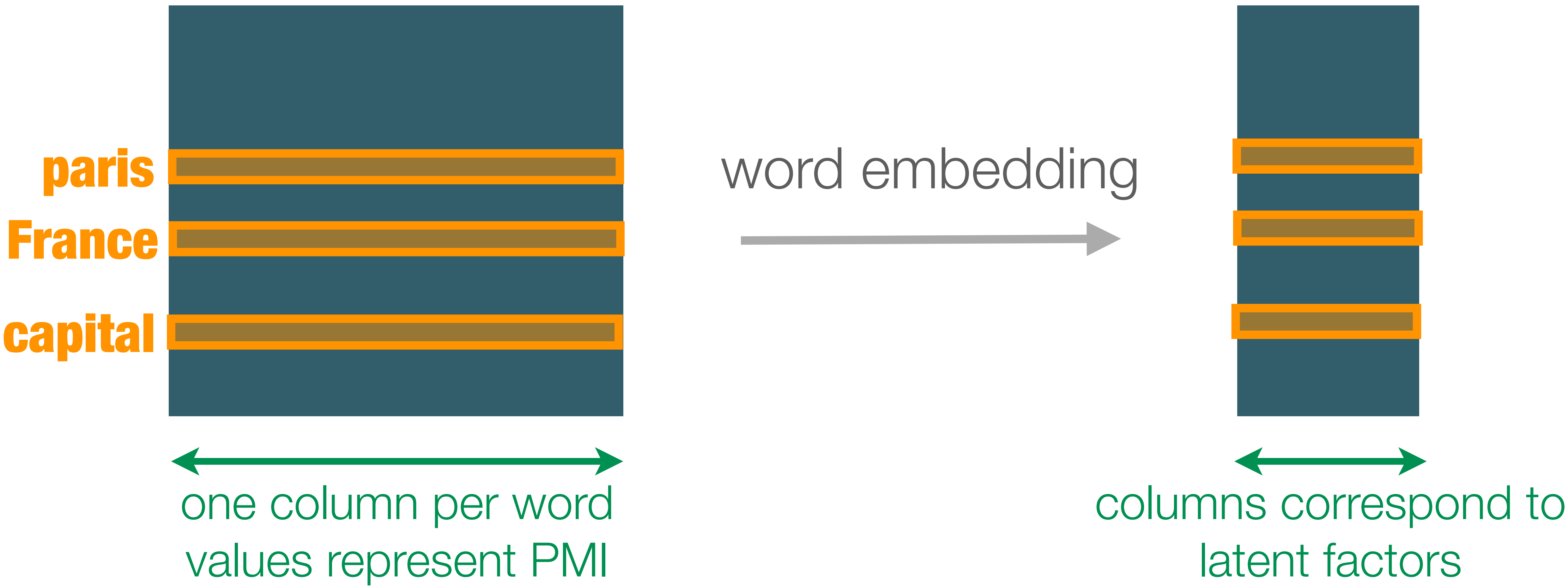
Carl Allen, Timothy M. Hospedales: Analogies Explained: Towards Understanding Word Embeddings. ICML 2019: 223-231

Why do vector differences model analogies?



Carl Allen, Timothy M. Hospedales: Analogies Explained: Towards Understanding Word Embeddings. ICML 2019: 223-231

Why do vector differences model analogies?



Carl Allen, Timothy M. Hospedales: Analogies Explained: Towards Understanding Word Embeddings. ICML 2019: 223-231

Abstract analogies

Query:		word:language
Candidates:	(1)	paint:portrait
	(2)	poetry:rhythm
	(3)	**note:music**
	(4)	tale:story
	(5)	week:year

Peter D. Turney: Measuring Semantic Similarity by Latent Relational Analysis. IJCAI 2005: 1136-1141

Abstract analogies

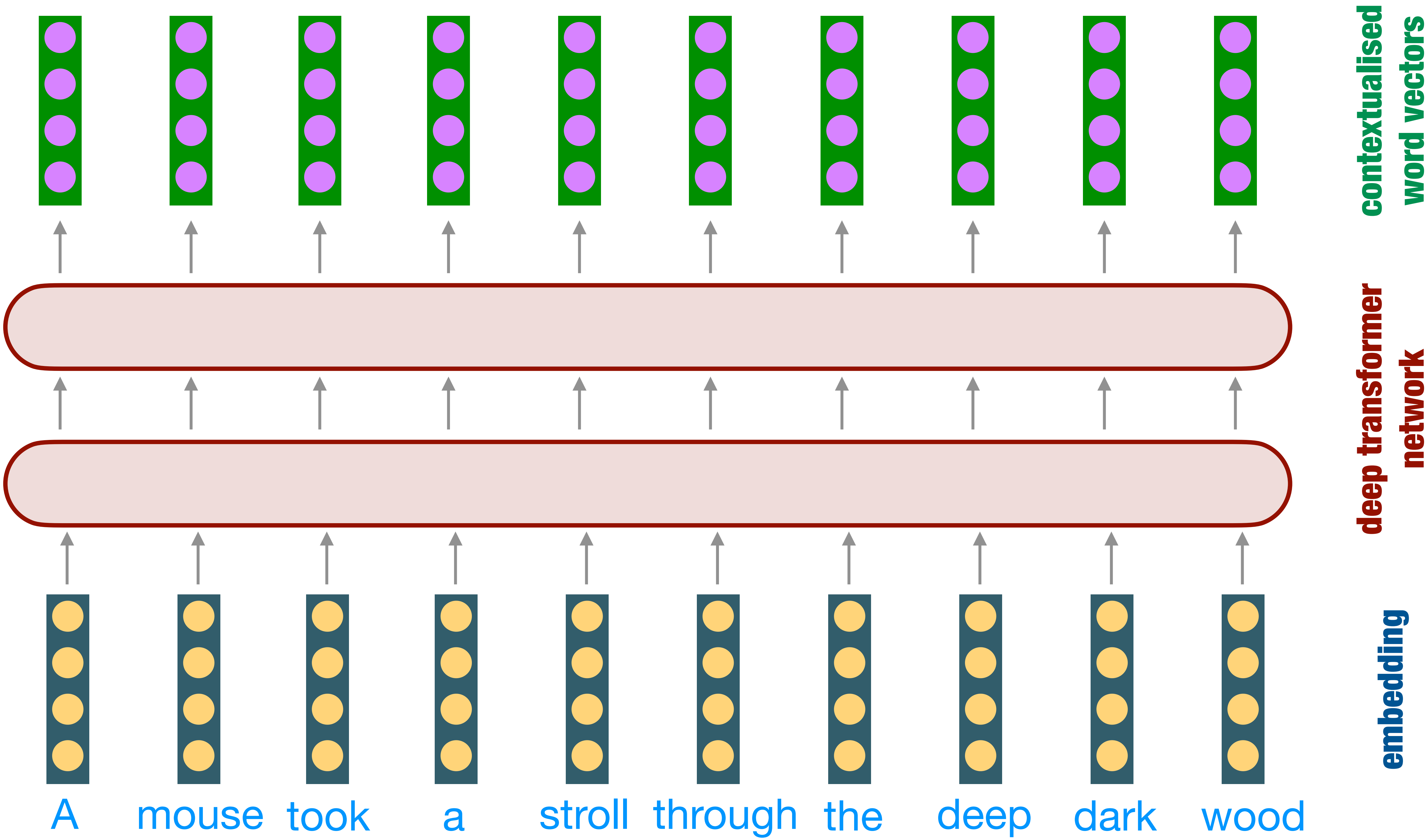
Query:		word:language
Candidates:	(1)	paint:portrait
	(2)	poetry:rhythm
	(3)	**note:music**
	(4)	tale:story
	(5)	week:year

Accuracy

- FastText: 49.7
- GloVe: 48.9
- Word2Vec: 42.8
- Latent Relational Analysis: 56.4

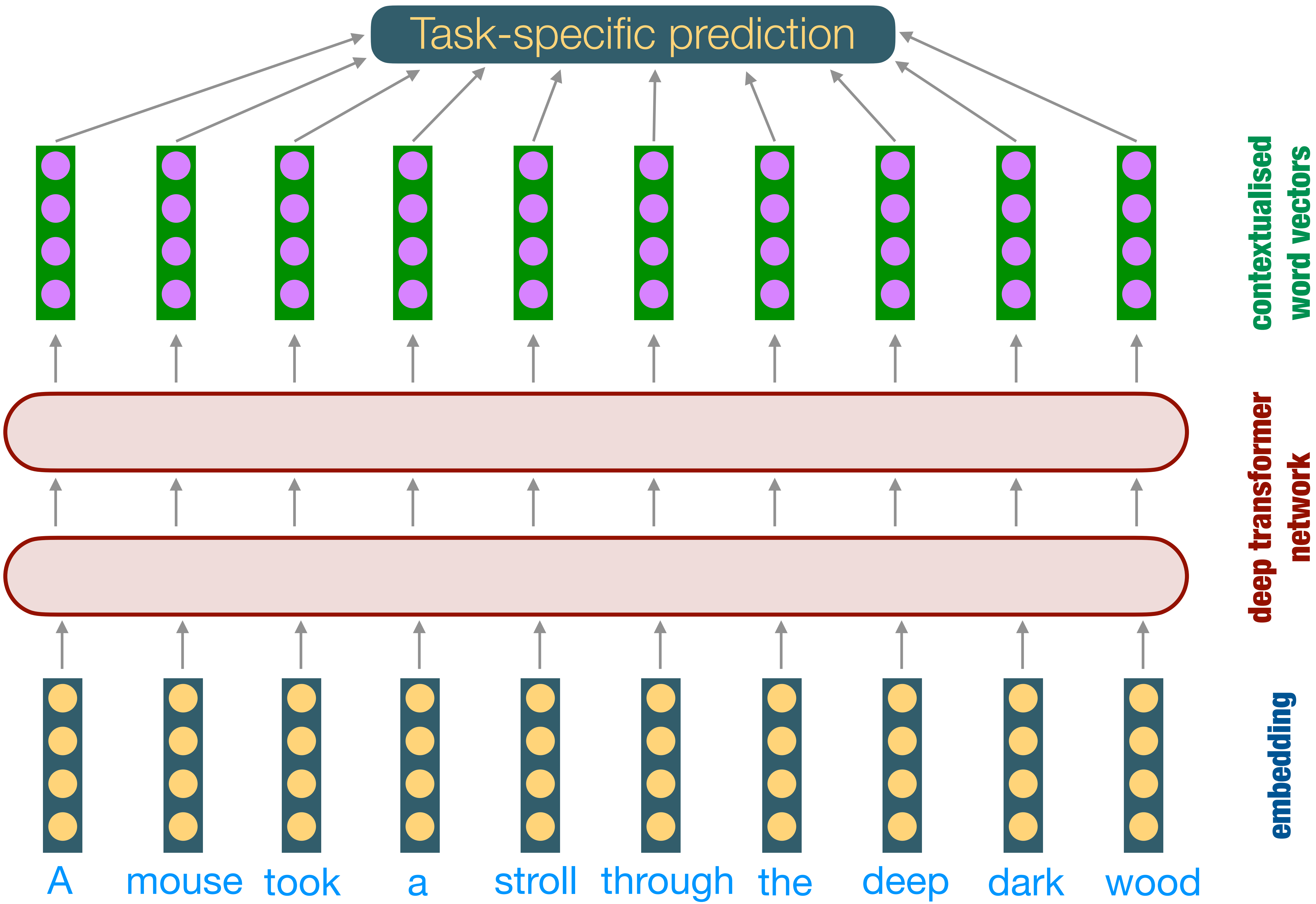
Peter D. Turney: Measuring Semantic Similarity by Latent Relational Analysis. IJCAI 2005: 1136-1141

Contextualised Language Models

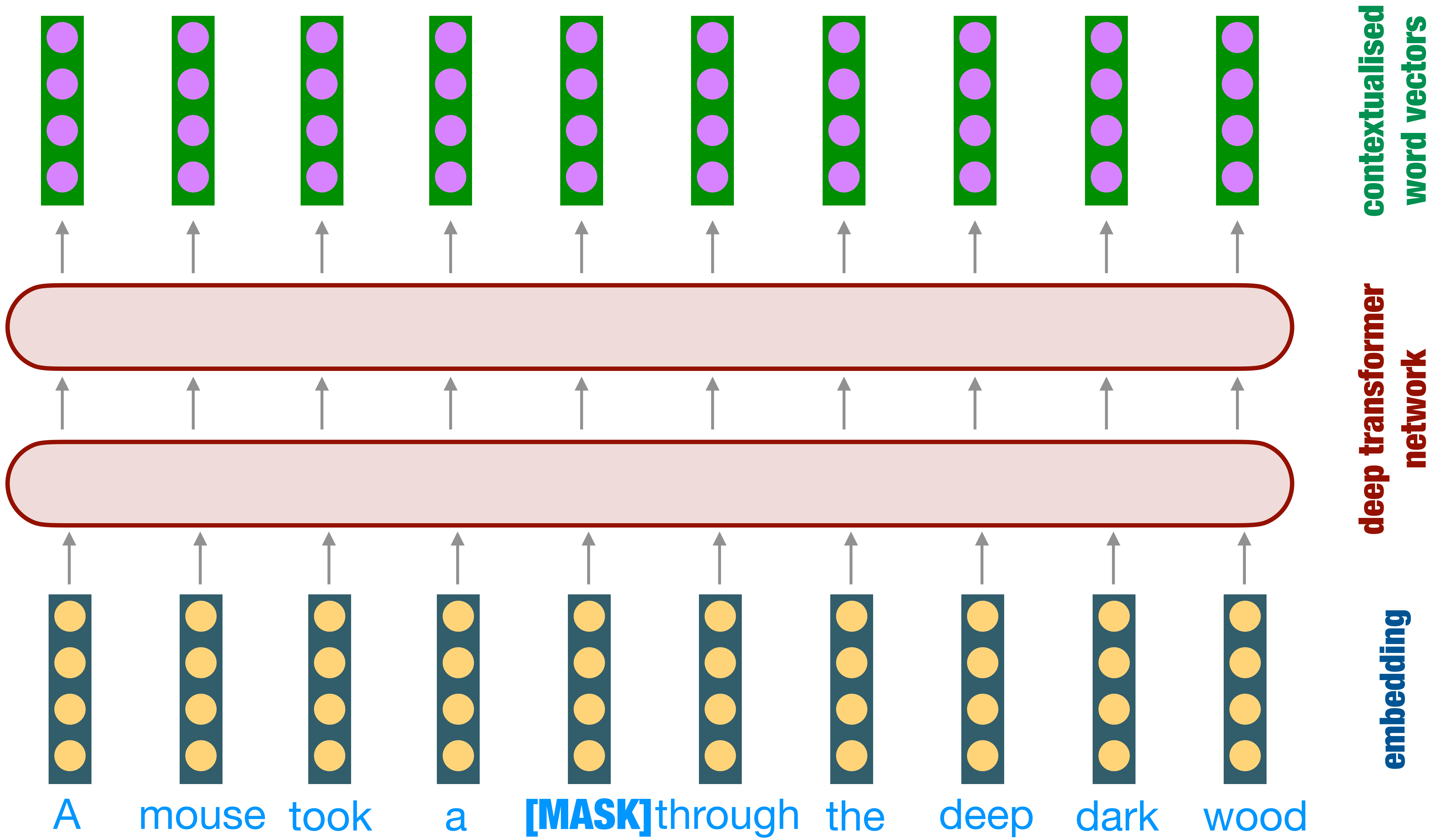


Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. NAACL-HLT 2019: 4171-4186

Contextualised Language Models

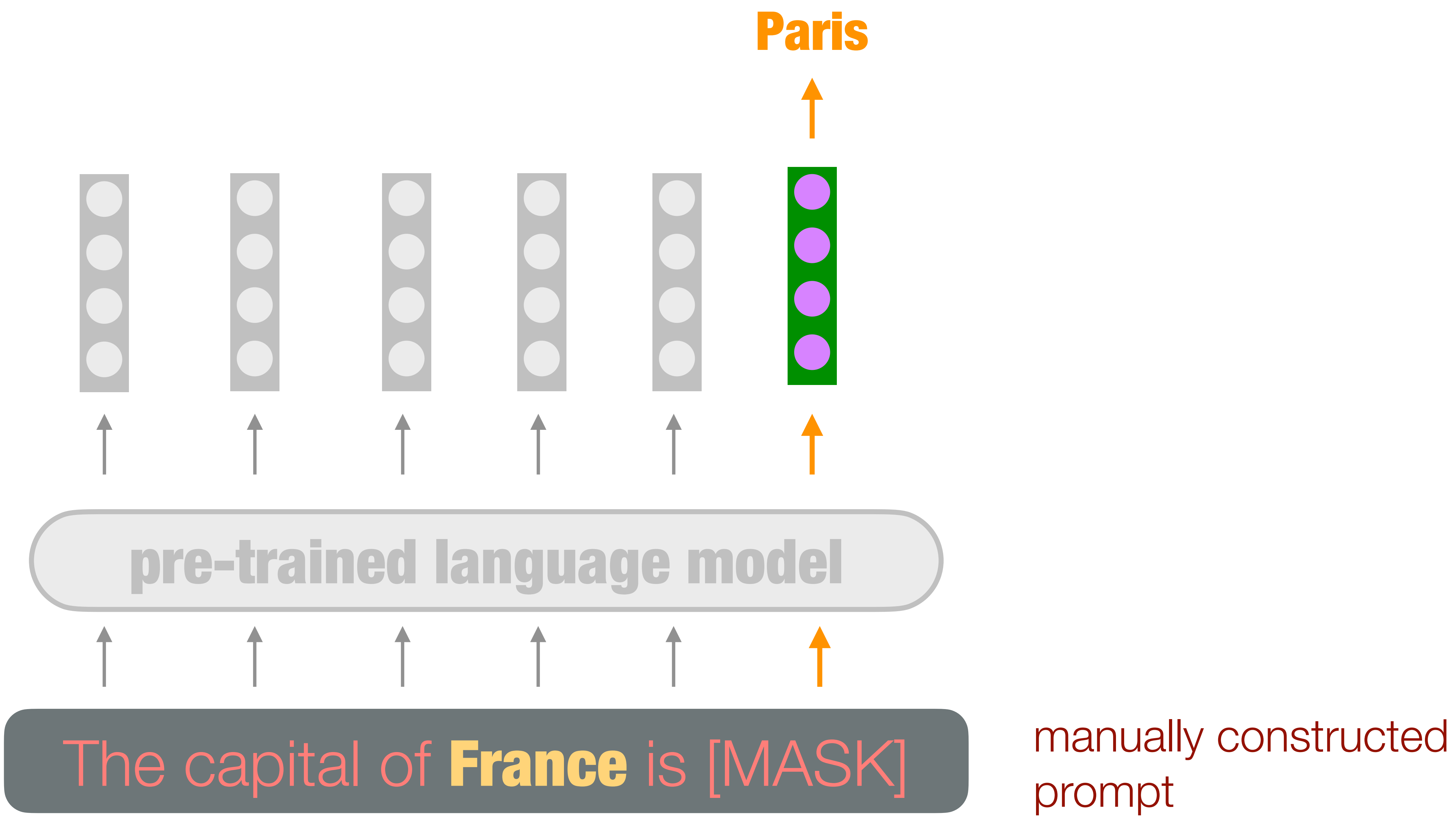


Contextualised Language Models



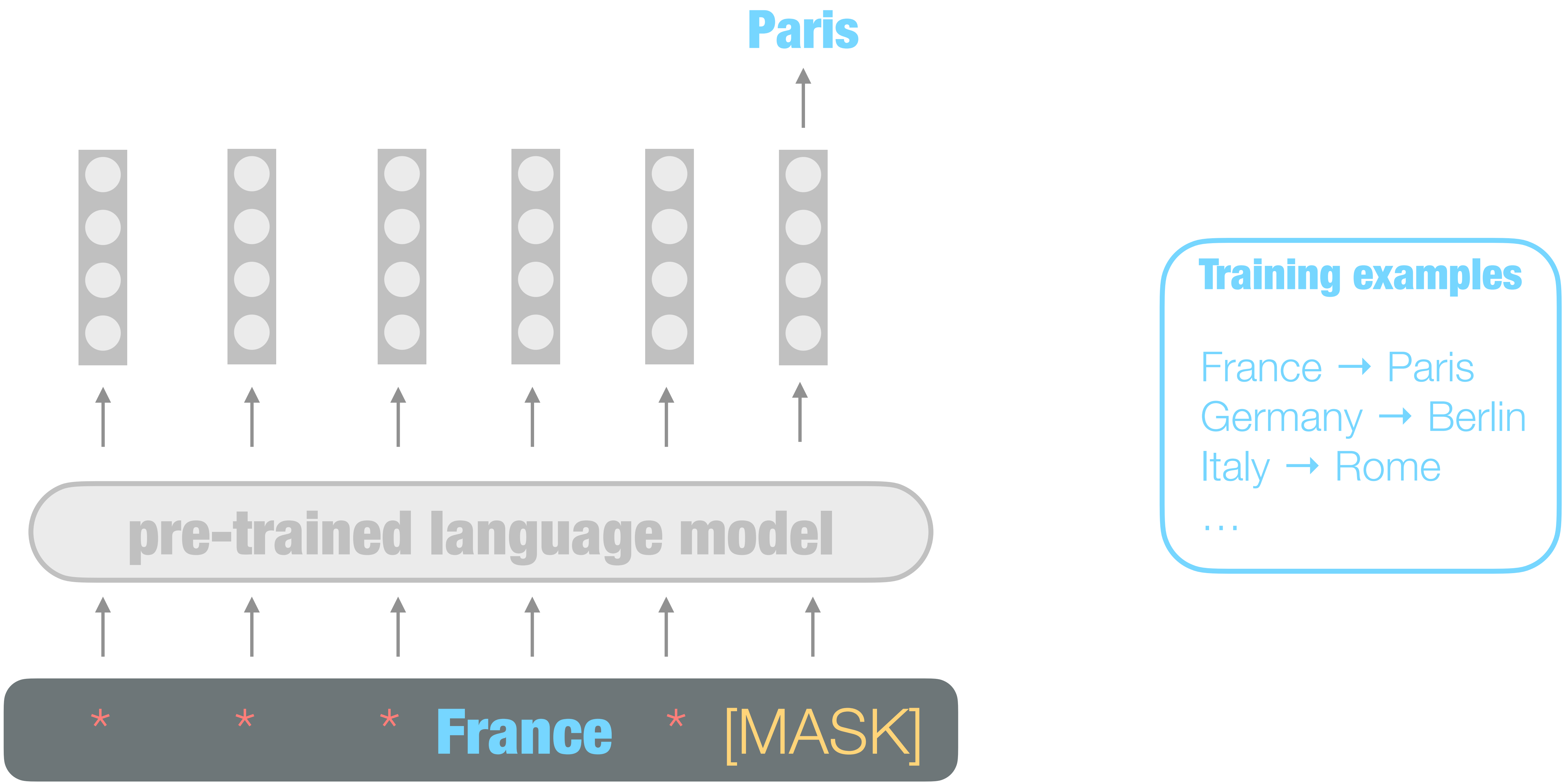
Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. NAACL-HLT 2019: 4171-4186

Language Models as Knowledge Bases



Fabio Petroni, Tim Rocktäschel, Sebastian Riedel, Patrick S. H. Lewis, Anton Bakhtin, Yuxiang Wu, Alexander H. Miller: Language Models as Knowledge Bases? EMNLP/IJCNLP (1) 2019: 2463-2473

Prompt Engineering

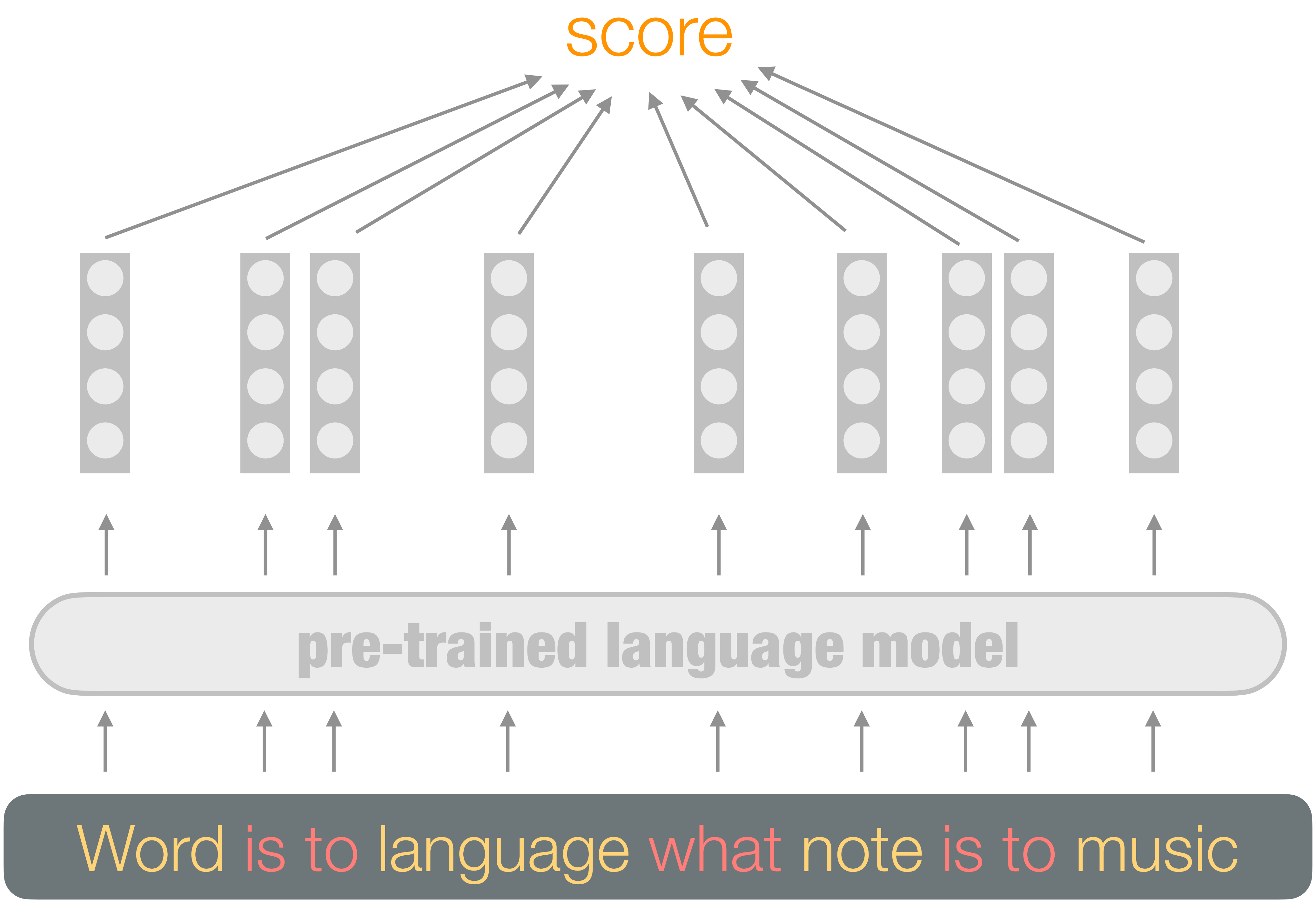


Taylor Shin, Yasaman Razeghi, Robert L. Logan IV, Eric Wallace, Sameer Singh: AutoPrompt: Eliciting Knowledge from Language Models with Automatically Generated Prompts. EMNLP (1) 2020: 4222-4235

BERT is to NLP what AlexNet is to CV: Can Pre-Trained Language Models Identify Analogies?

Asahi Ushio, Luis Espinosa-Anke, Steven Schockaert, Jose Camacho-Collados
Cardiff NLP, School of Computer Science and Informatics
Cardiff University, United Kingdom
{UshioA,Espinosa-AnkeL,SchockaertS1,CamachoColladosJ}@cardiff.ac.uk

Can language models recognise analogies?



Scoring functions: perplexity

How “fluent” are the following sentences:

word is to language what paint is to portrait
word is to language what poetry is to rhythm
word is to language what note is to music
word is to language what tale is to story
word is to language what week is to year

$$exp\left(-\sum_{j=1}^{m}\log P(x_j \mid x_{j-1})\right)$$

Scoring functions: PMI-based

How much more likely is “music” as the prediction in:

word is to language what note is to <u>[MASK]</u>

compared to:

word is to language what [MASK] is to <u>[MASK]</u>

$$\log P(t_i \mid h_i, h_q, t_q) - \alpha \log P(t_i \mid h_q, t_q)$$

Scoring functions: PMI-based

Compare the probability of the joint prediction (note,music) in:

word is to language what [MASK] is to [MASK]

to the probabilities of the individual predictions of note and music, respectively in:

word is to language what [MASK] is to [MASK]

word is to language what [MASK] is to [MASK]

$$\log P(t_i, h_i \mid h_q, t_q) - \alpha_t \log P(t_i \mid h_q, t_q) - \alpha_h \log P(h_i \mid h_q, t_q)$$

Results

Automatically learned prompt, optimised scoring function

	Model	Score	Tuned	SAT	U2	U4	Google	BATS	Avg
LM	BERT	s_{PPL}		32.9	32.9	34.0	80.8	61.5	48.4
			✓	39.8	41.7	41.0	86.8	67.9	55.4
		s_{PMI}		27.0	32.0	31.2	74.0	59.1	44.7
			✓	40.4	42.5	27.8	87.0	68.1	53.2
		s_{mPPL}	✓	41.8	44.7	41.2	88.8	67.9	56.9
	GPT-2	s_{PPL}		35.9	41.2	44.9	80.4	63.5	53.2
			✓	50.4	48.7	51.2	93.2	75.9	63.9
		s_{PMI}		34.4	44.7	43.3	62.8	62.8	49.6
			✓	51.0	37.7	50.5	91.0	79.8	62.0
		s_{mPPL}	✓	**56.7**	50.9	49.5	95.2	**81.2**	66.7
	RoBERTa	s_{PPL}		42.4	49.1	49.1	90.8	69.7	60.2
			✓	53.7	57.0	55.8	93.6	80.5	68.1
		s_{PMI}		35.9	42.5	44.0	60.8	60.8	48.8
			✓	51.3	49.1	38.7	92.4	77.2	61.7
		s_{mPPL}	✓	53.4	**58.3**	**57.4**	93.6	78.4	**68.2**
WE	FastText	-		47.8	43.0	40.7	**96.6**	72.0	60.0
	GloVe	-		47.8	46.5	39.8	96.0	68.7	59.8
	Word2vec	-		41.8	40.4	39.6	93.2	63.8	55.8
Base	PMI	-		23.3	32.9	39.1	57.4	42.7	39.1
	Random	-		20.0	23.6	24.2	25.0	25.0	23.6

Results

				abstract analogies			encyclopaedic and morphological knowledge		
	Model	Score	Tuned	SAT	U2	U4	Google	BATS	Avg
LM	BERT	s_{PPL}		32.9	32.9	34.0	80.8	61.5	48.4
			✓	39.8	41.7	41.0	86.8	67.9	55.4
		s_{PMI}		27.0	32.0	31.2	74.0	59.1	44.7
			✓	40.4	42.5	27.8	87.0	68.1	53.2
		s_{mPPL}	✓	41.8	44.7	41.2	88.8	67.9	56.9
	GPT-2	s_{PPL}		35.9	41.2	44.9	80.4	63.5	53.2
			✓	50.4	48.7	51.2	93.2	75.9	63.9
		s_{PMI}		34.4	44.7	43.3	62.8	62.8	49.6
			✓	51.0	37.7	50.5	91.0	79.8	62.0
		s_{mPPL}	✓	**56.7**	50.9	49.5	95.2	**81.2**	66.7
	RoBERTa	s_{PPL}		42.4	49.1	49.1	90.8	69.7	60.2
			✓	53.7	57.0	55.8	93.6	80.5	68.1
		s_{PMI}		35.9	42.5	44.0	60.8	60.8	48.8
			✓	51.3	49.1	38.7	92.4	77.2	61.7
		s_{mPPL}	✓	53.4	**58.3**	**57.4**	93.6	78.4	**68.2**
WE	FastText	-		47.8	43.0	40.7	**96.6**	72.0	60.0
	GloVe	-		47.8	46.5	39.8	96.0	68.7	59.8
	Word2vec	-		41.8	40.4	39.6	93.2	63.8	55.8
Base	PMI	-		23.3	32.9	39.1	57.4	42.7	39.1
	Random	-		20.0	23.6	24.2	25.0	25.0	23.6

Results

	Model	Score	Tuned	SAT	U2	U4	Google	BATS	Avg
LM	BERT	s_{PPL}		32.9	32.9	34.0	80.8	61.5	48.4
			✓	39.8	41.7	41.0	86.8	67.9	55.4
		s_{PMI}		27.0	32.0	31.2	74.0	59.1	44.7
			✓	40.4	42.5	27.8	87.0	68.1	53.2
		s_{mPPL}	✓	41.8	44.7	41.2	88.8	67.9	56.9
	GPT-2	s_{PPL}		35.9	41.2	44.9	80.4	63.5	53.2
			✓	50.4	48.7	51.2	93.2	75.9	63.9
		s_{PMI}		34.4	44.7	43.3	62.8	62.8	49.6
			✓	51.0	37.7	50.5	91.0	79.8	62.0
		s_{mPPL}	✓	**56.7**	50.9	49.5	95.2	**81.2**	66.7
	RoBERTa	s_{PPL}		42.4	49.1	49.1	90.8	69.7	60.2
			✓	53.7	57.0	55.8	93.6	80.5	68.1
		s_{PMI}		35.9	42.5	44.0	60.8	60.8	48.8
			✓	51.3	49.1	38.7	92.4	77.2	61.7
		s_{mPPL}	✓	53.4	**58.3**	**57.4**	93.6	78.4	**68.2**
WE	FastText	-		47.8	43.0	40.7	**96.6**	72.0	60.0
	GloVe	-		47.8	46.5	39.8	96.0	68.7	59.8
	Word2vec	-		41.8	40.4	39.6	93.2	63.8	55.8
Base	PMI	-		23.3	32.9	39.1	57.4	42.7	39.1
	Random	-		20.0	23.6	24.2	25.0	25.0	23.6

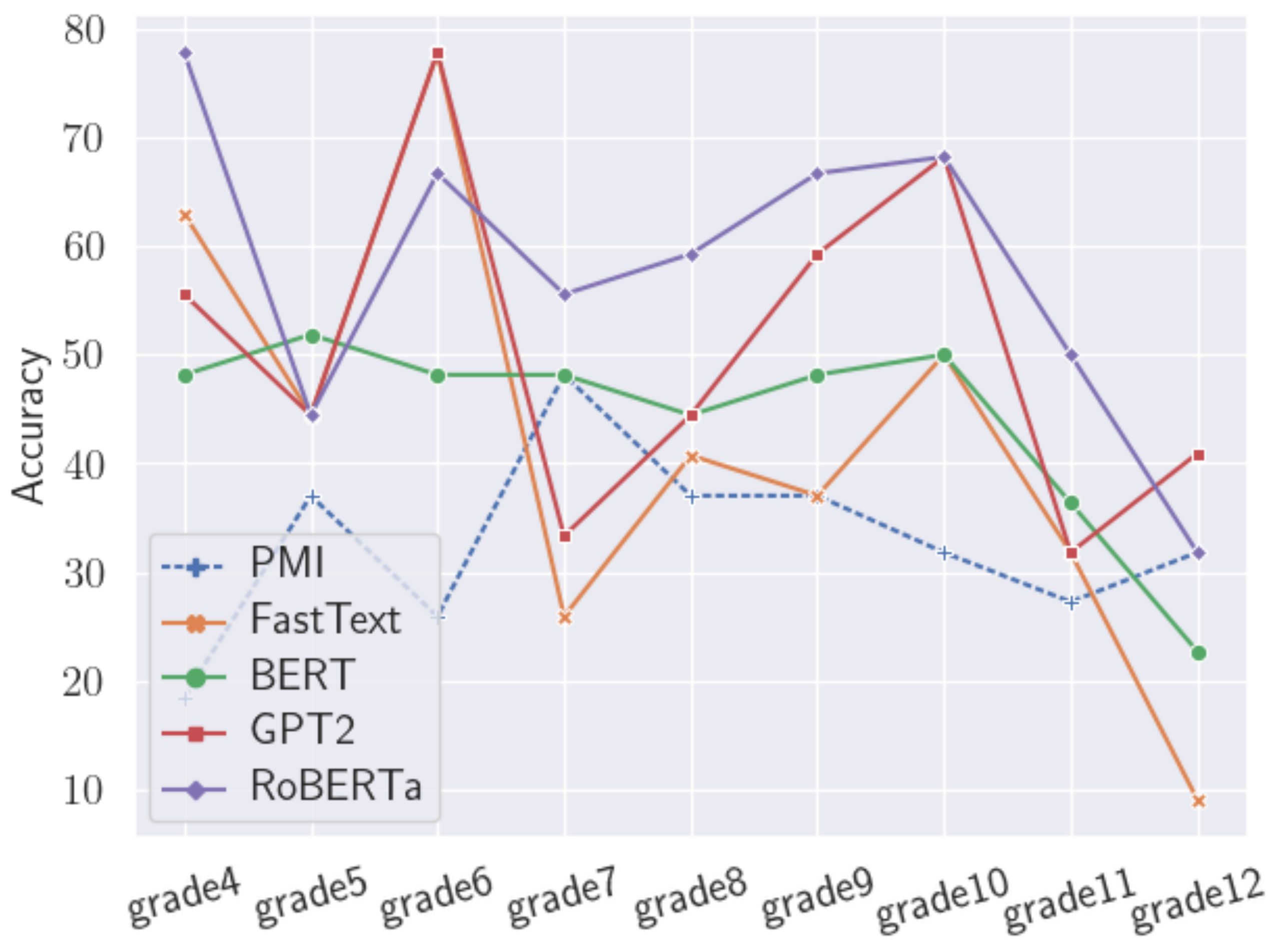
Results

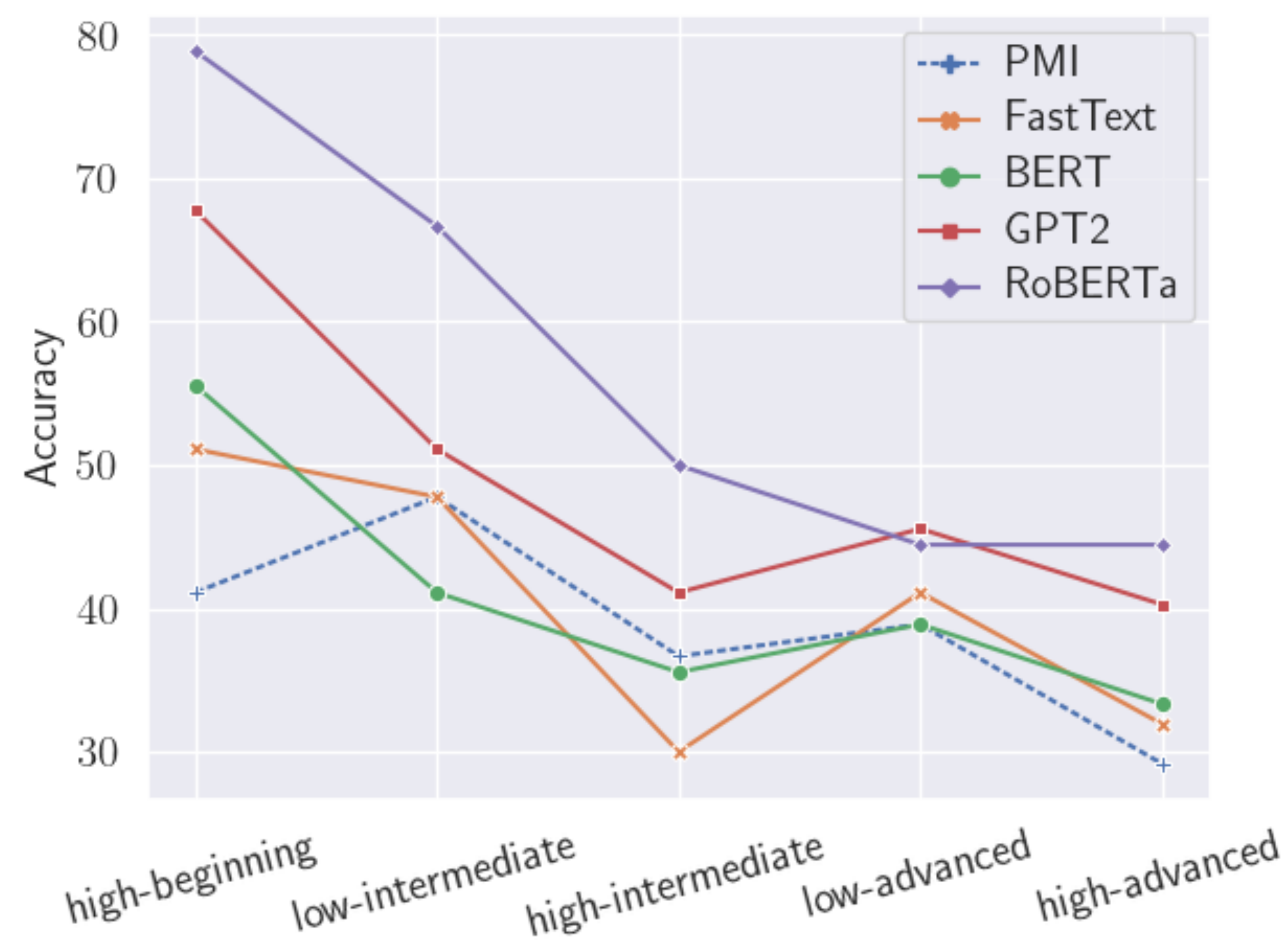
	Model	Score	Tuned	SAT	U2	U4	Google	BATS	Avg
LM	BERT	s_{PPL}		32.9	32.9	34.0	80.8	61.5	48.4
			✓	39.8	41.7	41.0	86.8	67.9	55.4
		s_{PMI}		27.0	32.0	31.2	74.0	59.1	44.7
			✓	40.4	42.5	27.8	87.0	68.1	53.2
		s_{mPPL}	✓	41.8	44.7	41.2	88.8	67.9	56.9
	GPT-2	s_{PPL}		35.9	41.2	44.9	80.4	63.5	53.2
			✓	50.4	48.7	51.2	93.2	75.9	63.9
		s_{PMI}		34.4	44.7	43.3	62.8	62.8	49.6
			✓	51.0	37.7	50.5	91.0	79.8	62.0
		s_{mPPL}	✓	**56.7**	50.9	49.5	95.2	**81.2**	66.7
	RoBERTa	s_{PPL}		42.4	49.1	49.1	90.8	69.7	60.2
			✓	53.7	57.0	55.8	93.6	80.5	68.1
		s_{PMI}		35.9	42.5	44.0	60.8	60.8	48.8
			✓	51.3	49.1	38.7	92.4	77.2	61.7
		s_{mPPL}	✓	53.4	**58.3**	**57.4**	93.6	78.4	**68.2**
WE	FastText	-		47.8	43.0	40.7	**96.6**	72.0	60.0
	GloVe	-		47.8	46.5	39.8	96.0	68.7	59.8
	Word2vec	-		41.8	40.4	39.6	93.2	63.8	55.8
Base	PMI	-		23.3	32.9	39.1	57.4	42.7	39.1
	Random	-		20.0	23.6	24.2	25.0	25.0	23.6

Results

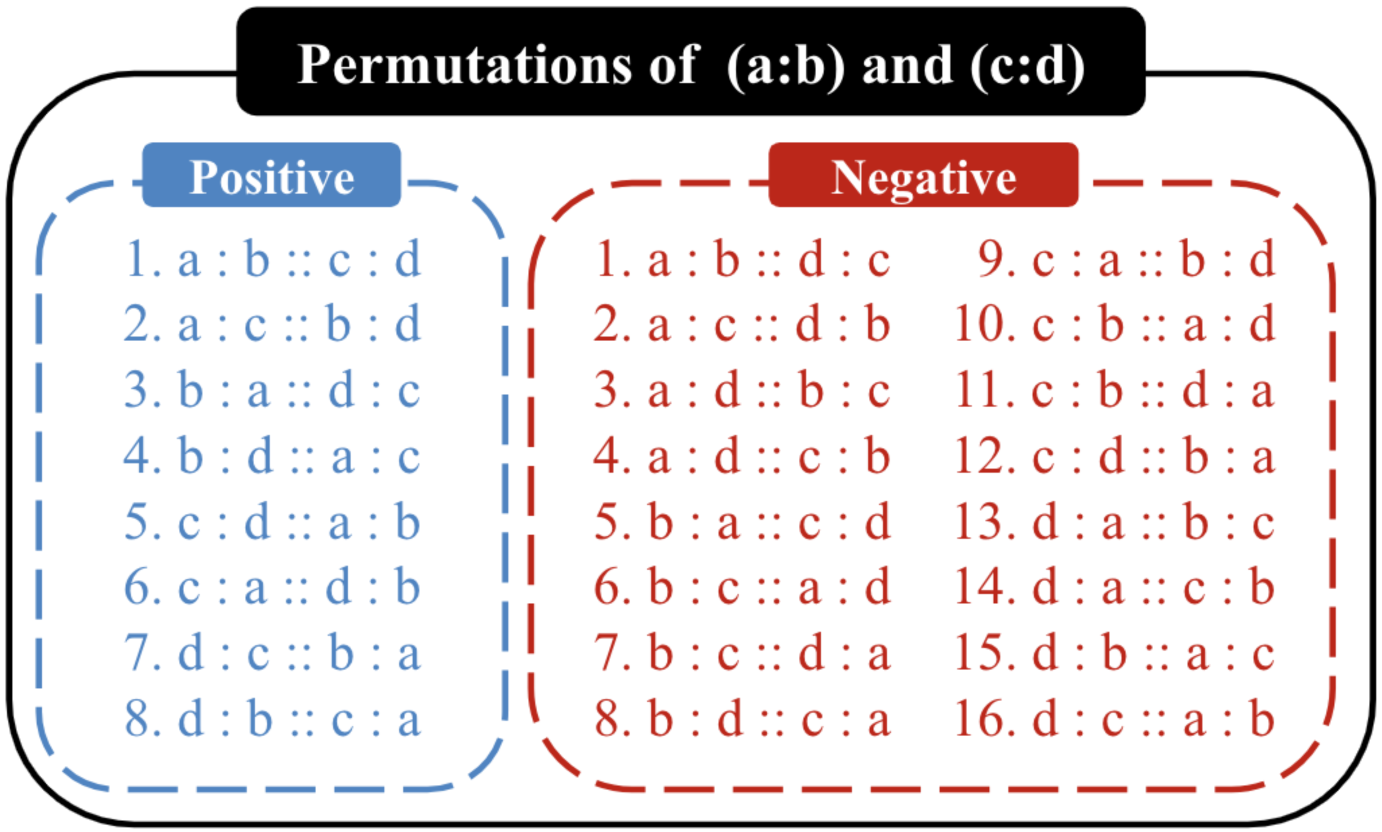
	Model	Score	Tuned	Accuracy
LM	BERT	s_{PPL}		32.6
			✓	40.4*
		s_{PMI}		26.8
			✓	41.2*
		s_{mPPL}	✓	42.8*
	GPT-2	s_{PPL}		41.4
			✓	56.2*
		s_{PMI}		34.7
			✓	56.8*
		s_{mPPL}	✓	57.8*
	RoBERTa	s_{PPL}		49.6
			✓	55.8*
		s_{PMI}		42.5
			✓	54.0*
		s_{mPPL}	✓	55.8*
	GPT-3	*Zero-shot*		*53.7*
		Few-shot	✓	*65.2**
-	LRA	-		*56.4*
WE	FastText	-		49.7
	GloVe	-		48.9
	Word2vec	-		42.8
Base	PMI	-		23.3
	Random	-		20.0

Easier for humans = easier for LMs?

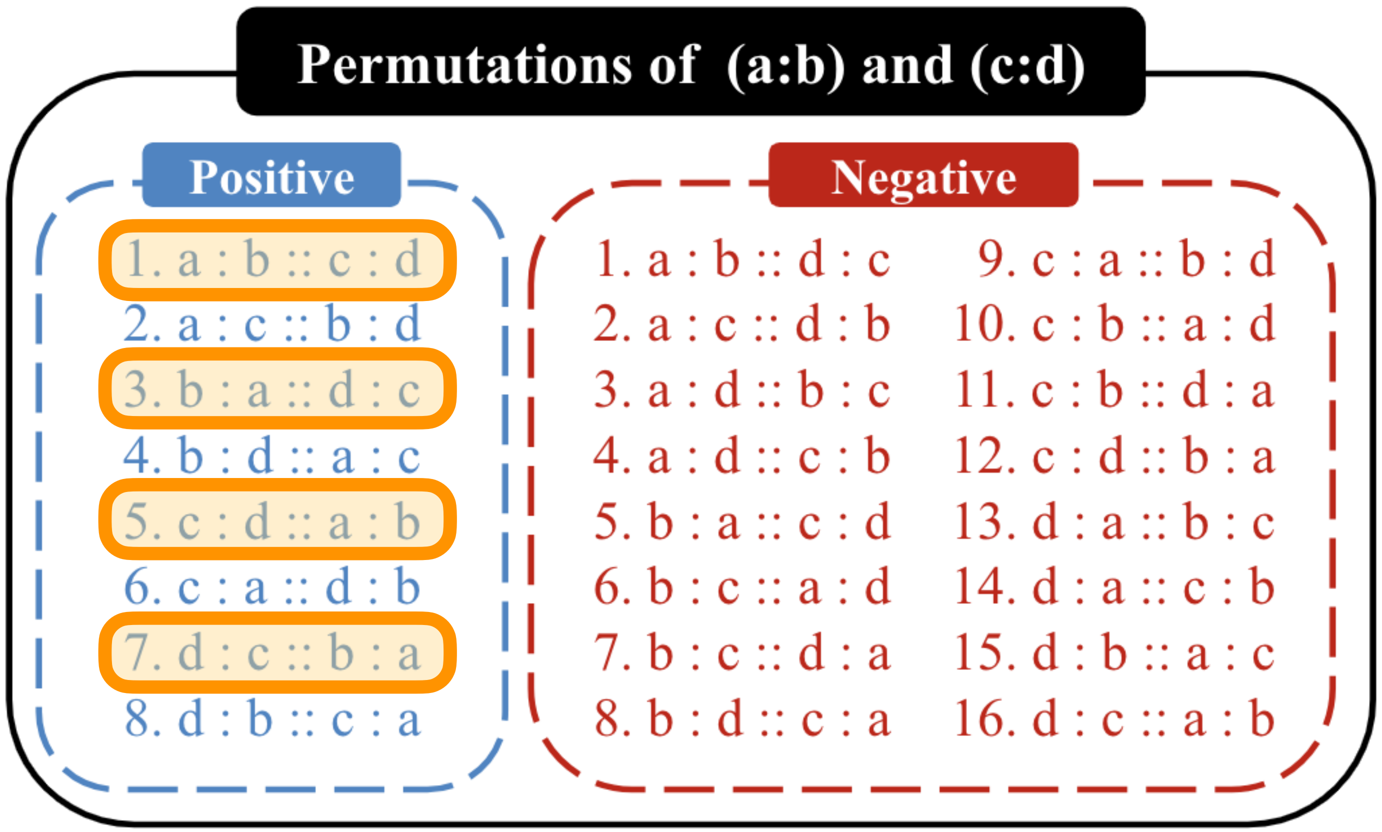


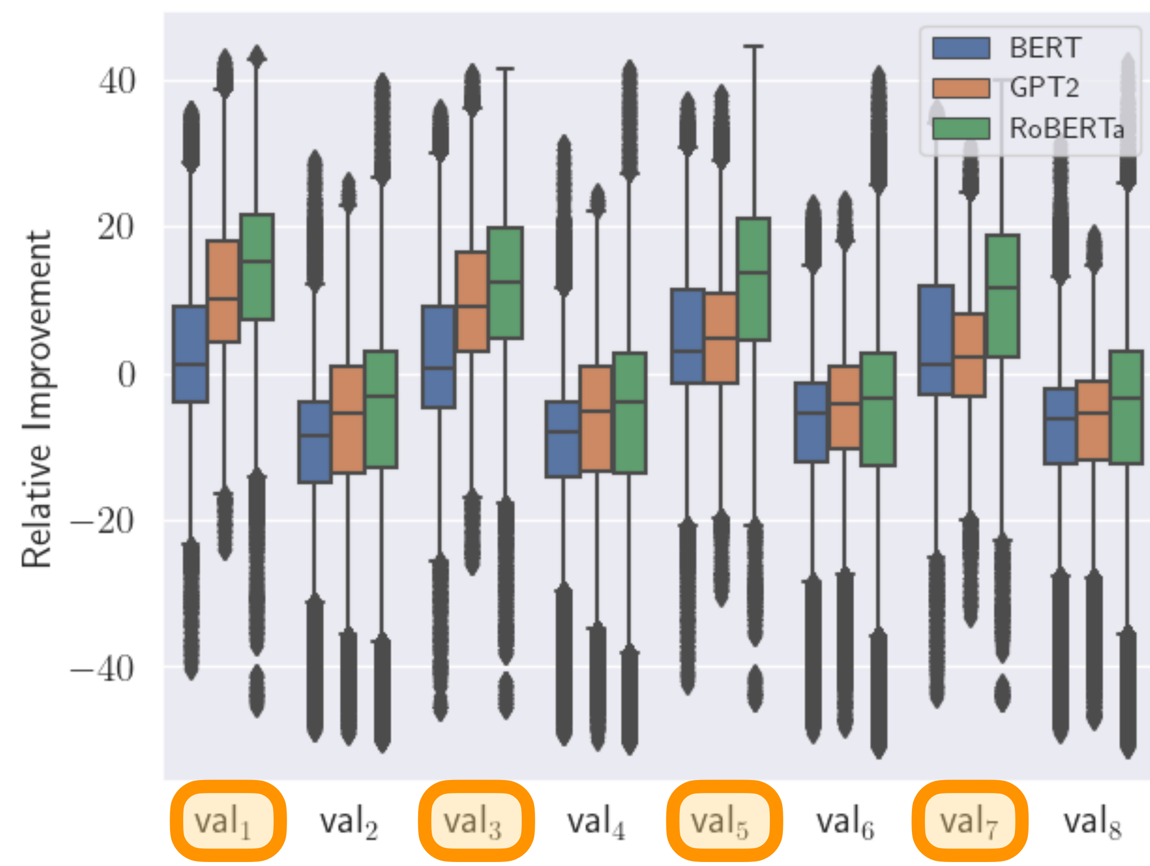


Are the results robust under permutations?



Are the results robust under permutations?

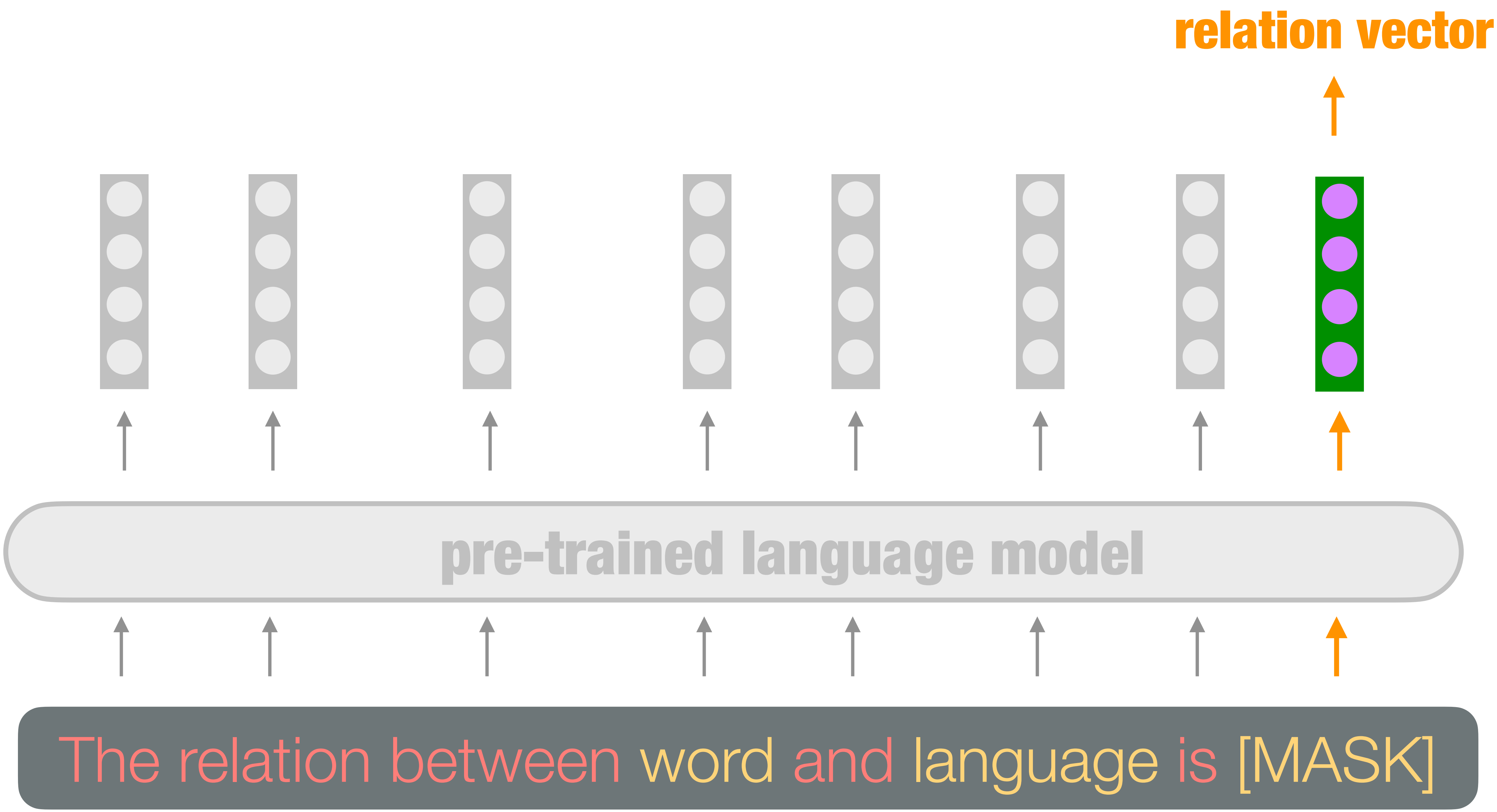




Distilling Relation Embeddings from Pre-trained Language Models

Asahi Ushio and **Jose Camacho-Collados** and **Steven Schockaert**
Cardiff NLP, School of Computer Science and Informatics
Cardiff University, United Kingdom
{UshioA,CamachoColladosJ,SchockaertS1}@cardiff.ac.uk

Learning relation vectors



Fabio Petroni, Tim Rocktäschel, Sebastian Riedel, Patrick S. H. Lewis, Anton Bakhtin, Yuxiang Wu, Alexander H. Miller: Language Models as Knowledge Bases? EMNLP/IJCNLP (1) 2019: 2463-2473

Fine-tune BERT on SemEval-2012 Task 2 data

Subcategory	Relation name	Relation schema	Paradigms	Responses
8(e)	AGENT:GOAL	"Y is the goal of X"	pilgrim:shrine assassin:death climber:peak	patient:health runner:finish astronaut:space
5(e)	OBJECT:TYPICAL ACTION	"an X will typically Y"	glass:break soldier:fight juggernaut:crush	ice:melt lion:roar knife:stab
4(h)	DEFECTIVE	"an X is is a defect in Y"	fallacy:logic astigmatism:sight limp:walk	pimple:skin ignorance:learning tumor:body

David Jurgens, Saif Mohammad, Peter D. Turney, Keith J. Holyoak: SemEval-2012 Task 2: Measuring Degrees of Relational Similarity. SemEval@NAACL-HLT 2012: 356-364

Fine-tune BERT on SemEval-2012 Task 2 data

66.0 "fire:hot"
59.6 "villain:evil"
53.8 "water:wet"
43.1 "tycoon:wealthy"
42.3 "snow:cold"
35.3 "candy:sweet"
32.0 "professor:intellectual"
30.0 "steel:strong"
30.0 "novice:inexperience"
…
-45.1 "prince:charming"
-46.0 "heat:fire"
-52.0 "lipstick:red"
-56.9 "fizzy:pop"
-60.0 "man:tall"
-72.5 "flimsy:paper"
-72.5 "tall:man"
-76.5 "intellectual:professor"

56.0 "loss:grief"
48.0 "injury:pain"
44.0 "disease:sickness"
42.0 "explosion:damage"
41.2 "accident:damage"
34.5 "germs:sickness"
30.0 "bath:cleanliness"
26.0 "exercise:fitness"
22.0 "tragedy:tears"
…
-22.0 "digging:hole"
-24.0 "sow:germinate"
-28.0 "yelling:anger"
-42.0 "headache:stress"
-48.0 "learning:study"
-62.0 "response:stimulus"
-66.0 "boredom:repetition"
-74.0 "sweat:run"

…

David Jurgens, Saif Mohammad, Peter D. Turney, Keith J. Holyoak: SemEval-2012 Task 2: Measuring Degrees of Relational Similarity. SemEval@NAACL-HLT 2012: 356-364

Fine-tune BERT on SemEval-2012 Task 2 data

should be similar

66.0 "fire:hot"	56.0 "loss:grief"
59.6 "villain:evil"	48.0 "injury:pain"
53.8 "water:wet"	44.0 "disease:sickness"
43.1 "tycoon:wealthy"	42.0 "explosion:damage"
42.3 "snow:cold"	41.2 "accident:damage"
35.3 "candy:sweet"	34.5 "germs:sickness"
32.0 "professor:intellectual"	30.0 "bath:cleanliness"
30.0 "steel:strong"	26.0 "exercise:fitness"
30.0 "novice:inexperience"	22.0 "tragedy:tears"
…	…
-45.1 "prince:charming"	-22.0 "digging:hole"
-46.0 "heat:fire"	-24.0 "sow:germinate"
-52.0 "lipstick:red"	-28.0 "yelling:anger"
-56.9 "fizzy:pop"	-42.0 "headache:stress"
-60.0 "man:tall"	-48.0 "learning:study"
-72.5 "flimsy:paper"	-62.0 "response:stimulus"
-72.5 "tall:man"	-66.0 "boredom:repetition"
-76.5 "intellectual:professor"	-74.0 "sweat:run"

…

David Jurgens, Saif Mohammad, Peter D. Turney, Keith J. Holyoak: SemEval-2012 Task 2: Measuring Degrees of Relational Similarity. SemEval@NAACL-HLT 2012: 356-364

Fine-tune BERT on SemEval-2012 Task 2 data

should be dissimilar

66.0 "fire:hot"
59.6 "villain:evil"
53.8 "water:wet"
43.1 "tycoon:wealthy"
42.3 "snow:cold"
35.3 "candy:sweet"
32.0 "professor:intellectual"
30.0 "steel:strong"
30.0 "novice:inexperience"
…
-45.1 "prince:charming"
-46.0 "heat:fire"
-52.0 "lipstick:red"
-56.9 "fizzy:pop"
-60.0 "man:tall"
-72.5 "flimsy:paper"
-72.5 "tall:man"
-76.5 "intellectual:professor"

56.0 "loss:grief"
48.0 "injury:pain"
44.0 "disease:sickness"
42.0 "explosion:damage"
41.2 "accident:damage"
34.5 "germs:sickness"
30.0 "bath:cleanliness"
26.0 "exercise:fitness"
22.0 "tragedy:tears"
…
-22.0 "digging:hole"
-24.0 "sow:germinate"
-28.0 "yelling:anger"
-42.0 "headache:stress"
-48.0 "learning:study"
-62.0 "response:stimulus"
-66.0 "boredom:repetition"
-74.0 "sweat:run"

…

David Jurgens, Saif Mohammad, Peter D. Turney, Keith J. Holyoak: SemEval-2012 Task 2: Measuring Degrees of Relational Similarity. SemEval@NAACL-HLT 2012: 356-364

Fine-tune BERT on SemEval-2012 Task 2 data

should be dissimilar

66.0 "fire:hot"
59.6 "villain:evil"
53.8 "water:wet"
43.1 "tycoon:wealthy"
42.3 "snow:cold"
35.3 "candy:sweet"
32.0 "professor:intellectual"
30.0 "steel:strong"
30.0 "novice:inexperience"
…
-45.1 "prince:charming"
-46.0 "heat:fire"
-52.0 "lipstick:red"
-56.9 "fizzy:pop"
-60.0 "man:tall"
-72.5 "flimsy:paper"
-72.5 "tall:man"
-76.5 "intellectual:professor"

56.0 "loss:grief"
48.0 "injury:pain"
44.0 "disease:sickness"
42.0 "explosion:damage"
41.2 "accident:damage"
34.5 "germs:sickness"
30.0 "bath:cleanliness"
26.0 "exercise:fitness"
22.0 "tragedy:tears"
…
-22.0 "digging:hole"
-24.0 "sow:germinate"
-28.0 "yelling:anger"
-42.0 "headache:stress"
-48.0 "learning:study"
-62.0 "response:stimulus"
-66.0 "boredom:repetition"
-74.0 "sweat:run"

…

David Jurgens, Saif Mohammad, Peter D. Turney, Keith J. Holyoak: SemEval-2012 Task 2: Measuring Degrees of Relational Similarity. SemEval@NAACL-HLT 2012: 356-364

Training loss

Triplet loss

$$L_t = \max\left(0, \|\boldsymbol{x}_a - \boldsymbol{x}_p\| - \|\boldsymbol{x}_a - \boldsymbol{x}_n\| + \varepsilon\right)$$

relation vector for some word pair a

relation vector for a word pair p, similar to a

relation vector for a word pair n, not similar to a

David Jurgens, Saif Mohammad, Peter D. Turney, Keith J. Holyoak: SemEval-2012 Task 2: Measuring Degrees of Relational Similarity. SemEval@NAACL-HLT 2012: 356-364

Training loss

Triplet loss

$$L_t = \max\left(0, \|\boldsymbol{x}_a - \boldsymbol{x}_p\| - \|\boldsymbol{x}_a - \boldsymbol{x}_n\| + \varepsilon\right)$$

Classification loss

$$L_c = -\log(g(\boldsymbol{x}_a, \boldsymbol{x}_p)) - \log(1 - g(\boldsymbol{x}_a, \boldsymbol{x}_n))$$

$$g(\boldsymbol{u}, \boldsymbol{v}) = \mathrm{sigmoid}(W \cdot [\boldsymbol{u} \oplus \boldsymbol{v} \oplus |\boldsymbol{v} - \boldsymbol{u}|]^T)$$

David Jurgens, Saif Mohammad, Peter D. Turney, Keith J. Holyoak: SemEval-2012 Task 2: Measuring Degrees of Relational Similarity. SemEval@NAACL-HLT 2012: 356-364

Results

Model	**SAT†**	**SAT**	**U2**	**U4**	**Google**	**BATS**
Random	20.0	20.0	23.6	24.2	25.0	25.0
PMI	23.3	23.1	32.9	39.1	57.4	42.7
LRA	*56.4*	-	-	-	-	-
SuperSim	*54.8*	-	-	-	-	-
GPT-3 (zero)	*53.7*	-	-	-	-	-
GPT-3 (few)	*65.2**	-	-	-	-	-
RELATIVE	24.9	24.6	32.5	27.1	62.0	39.0
pair2vec	33.7	34.1	25.4	28.2	66.6	53.8
GloVe	48.9	47.8	46.5	39.8	96.0	68.7
FastText	49.7	47.8	43.0	40.7	**96.6**	72.0
Analogical Proportion Score						
· GPT-2	*41.4*	*35.9*	*41.2*	*44.9*	*80.4*	*63.5*
· BERT	*32.6*	*32.9*	*32.9*	*34.0*	*80.8*	*61.5*
· RoBERTa	*49.6*	*42.4*	*49.1*	*49.1*	*90.8*	*69.7*
Analogical Proportion Score (tuned)						
· GPT-2	*57.8**	*56.7**	*50.9**	*49.5**	*95.2**	*81.2**
· BERT	*42.8**	*41.8**	*44.7**	*41.2**	*88.8**	*67.9**
· RoBERTa	*55.8**	*53.4**	*58.3**	*57.4**	*93.6**	*78.4**
RelBERT						
· Manual	**69.5**	**70.6**	**66.2**	**65.3**	92.4	**78.8**
· AutoPrompt	61.0	62.3	61.4	63.0	88.2	74.6
· P-tuning	54.0	55.5	58.3	55.8	83.4	72.1

Results

Model		BLESS		CogALexV		EVALution		K&H+N		ROOT09	
		macro	micro	macro	micro	macro	micro	macro	micro	macro	micro
GloVe	*cat*	92.9	93.3	42.8	73.5	56.9	58.3	88.8	94.9	86.3	86.5
	cat+dot	**93.1**	93.7	51.9	79.2	55.9	57.3	89.6	95.1	88.8	89.0
	cat+dot+pair	91.8	92.6	56.4	81.1	58.1	59.6	89.4	95.7	89.2	89.4
	cat+dot+rel	91.1	92.0	53.2	79.2	58.4	58.6	89.3	94.9	89.3	89.4
	diff	91.0	91.5	39.2	70.8	55.6	56.9	87.0	94.4	85.9	86.3
	diff+dot	92.3	92.9	50.6	78.5	56.5	57.9	88.3	94.8	88.6	88.9
	diff+dot+pair	91.3	92.2	55.5	80.2	56.0	57.4	88.0	95.5	89.1	89.4
	diff+dot+rel	91.1	91.8	52.8	78.6	56.9	57.9	87.4	94.6	87.7	88.1
FastText	*cat*	92.4	92.9	40.7	72.4	56.4	57.9	88.1	93.8	85.7	85.5
	cat+dot	92.7	93.2	48.5	77.4	56.7	57.8	89.1	94.0	88.2	88.5
	cat+dot+pair	90.9	91.5	53.0	79.3	56.1	58.2	88.3	94.3	87.7	87.8
	cat+dot+rel	91.4	91.9	50.6	76.8	57.9	59.1	86.9	93.5	87.1	87.4
	diff	90.7	91.2	39.7	70.2	53.2	55.5	85.8	93.3	85.5	86.0
	diff+dot	92.3	92.9	49.1	77.8	55.2	57.4	86.5	93.6	88.5	88.9
	diff+dot+pair	90.0	90.8	53.9	79.0	55.8	57.8	86.6	94.2	87.7	88.1
	diff+dot+rel	90.6	91.3	53.6	78.2	57.1	58.0	86.3	93.4	86.9	87.4
RelBERT	Manual	91.7	92.1	**71.2**	**87.0**	68.4	69.6	88.0	96.2	**90.9**	**91.0**
	AutoPrompt	91.9	92.4	68.5	85.1	**69.5**	**70.5**	**91.3**	97.1	90.0	90.3
	P-tuning	91.3	91.8	67.8	84.9	69.1	70.2	88.5	96.3	89.8	89.9
SotA	LexNET	-	89.3	-	-	-	60.0	-	98.5	-	81.3
	SphereRE	-	**93.8**	-	-	-	62.0	-	**99.0**	-	86.1

Results

	BLESS	CogALex	EVAL	K&H+N	ROOT09
rand	93.7 (+0.3)	94.3 (-0.2)	-	97.9 (+0.2)	91.2 (-0.1)
mero	89.8 (+1.4)	72.9 (+2.7)	69.2 (+1.9)	74.5 (+5.4)	-
event	86.5 (-0.3)	-	-	-	-
hyp	94.1 (+0.8)	60.9 (-0.7)	61.7 (-1.5)	93.5 (+5.0)	83.0 (-0.4)
cohyp	96.4 (+0.3)	-	-	97.8 (+1.2)	97.4 (-0.5)
attr	92.6 (+0.3)	-	84.7 (+1.6)	-	-
poss	-	-	67.1 (-0.2)	-	-
ant	-	76.8 (-2.6)	81.3 (-0.9)	-	-
syn	-	49.9 (-0.6)	53.6 (+2.7)	-	-
macro	92.2 (+0.5)	71.0 (-0.2)	69.3 (+0.9)	90.9 (+2.9)	90.5 (-0.4)
micro	92.5 (+0.4)	86.9 (-0.1)	70.2 (+0.6)	97.2 (+1.0)	90.7 (-0.3)

Model was trained without any hypernymy training data

Examples

Category	Target	Nearest Neighbors RelBERT
Commonsense	barista:coffee	baker:bread, brewer:beer, bartender:cocktail, winemaker:wine, bartender:drink, baker:cake
	restaurant:waitress	restaurant:waiter, diner:waitress, bar:bartender, hospital:nurse, courthouse:clerk, office:clerk
	car:garage	car:pit, plane:hangar, auto:garage, baby:crib, yacht:harbour, aircraft:hangar
	ice:melt	snow:melt, glacier:melt, ice:drift, crust:melt, polar ice:melt, ice:thaw
	dolphin:swim	squid:swim, salmon:swim, shark:swim, fish:swim, horse:run, frog:leap
	flower:fragrant	orchid:fragrant, cluster:fragrant, jewel:precious, jewel:valuable, soil:permeable, vegetation:abundant
	coconut:milk	coconut:oil, goat:milk, grape:juice, palm:oil, olive:oil, camel:milk
	bag:plastic	bottle:plastic, bag:leather, container:plastic, box:plastic, jug:glass, bottle:glass
	duck:duckling	chicken:chick, pig:piglet, cat:kitten, ox:calf, butterfly:larvae, bear:cub
Gender	man:woman	men:women, male:female, father:mother, boy:girl, hero:heroine, king:queen
Antonymy	cooked:raw	raw:cooked, regulated:unregulated, sober:drunk, loaded:unloaded, armed:unarmed, published:unpublished
	normal:abnormal	ordinary:unusual, usual:unusual, acceptable:unacceptable, stable:unstable, rational:irrational, legal:illegal
Meronymy	helicopter:rotor	helicopter:rotor blades, helicopter:wing, bicycle:wheel, motorcycle:wheel, airplane:engine, plane:engine
	bat:wing	butterfly:wing, eagle:wing, angel:wing, cat:paw, lion:wings, fly:wing
	beer:alcohol	wine:alcohol, cider:alcohol, soda:sugar, beer:liquor, beer:malt, lager:alcoho
	oxygen:atmosphere	helium:atmosphere, hydrogen:atmosphere, nitrogen:atmosphere, methane:atmosphere, carbon:atmosphere
Hypernymy	chihuahua:dog	dachshund:dog, poodle:dog, terrier:dog, chinchilla:rodent, macaque:monkey, dalmatian:dog
	pelican:bird	toucan:bird, puffin:bird, egret:bird, peacock:bird, grouse:bird, pigeon:bird
	tennis:sport	hockey:sport, soccer:sport, volleyball:sport, cricket:sport, golf:sport, football:sport
Morphology	dog:dogs	cat:cats, horse:horses, pig:pigs, rat:rats, wolf:wolves, monkey:monkeys
	tall:tallest	strong:strongest, short:shortest, smart:smartest, weak:weakest, big:biggest, small:smallest
	spy:espionage	pirate:piracy, robber:robbery, lobbyist:lobbying, scout:scouting, terrorist:terrorism, witch:witchcraft

Examples

Category	Target	Nearest Neighbors RelBERT
Commonsense	barista:coffee	baker:bread, brewer:beer, bartender:cocktail, winemaker:wine, bartender:drink, baker:cake
	restaurant:waitress	restaurant:waiter, diner:waitress, bar:bartender, hospital:nurse, courthouse:clerk, office:clerk
	car:garage	car:pit, plane:hangar, auto:garage, baby:crib, yacht:harbour, aircraft:hangar
	ice:melt	snow:melt, glacier:melt, ice:drift, crust:melt, polar ice:melt, ice:thaw
	dolphin:swim	squid:swim, salmon:swim, shark:swim, fish:swim, horse:run, frog:leap
	flower:fragrant	orchid:fragrant, cluster:fragrant, jewel:precious, jewel:valuable, soil:permeable, vegetation:abundant
	coconut:milk	coconut:oil, goat:milk, grape:juice, palm:oil, olive:oil, camel:milk
	bag:plastic	bottle:plastic, bag:leather, container:plastic, box:plastic, jug:glass, bottle:glass
	duck:duckling	chicken:chick, pig:piglet, cat:kitten, ox:calf, butterfly:larvae, bear:cub
Gender	man:woman	men:women, male:female, father:mother, boy:girl, hero:heroine, king:queen
Antonymy	cooked:raw	raw:cooked, regulated:unregulated, sober:drunk, loaded:unloaded, armed:unarmed, published:unpublished
	normal:abnormal	ordinary:unusual, usual:unusual, acceptable:unacceptable, stable:unstable, rational:irrational, legal:illegal
Meronymy	helicopter:rotor	helicopter:rotor blades, helicopter:wing, bicycle:wheel, motorcycle:wheel, airplane:engine, plane:engine
	bat:wing	butterfly:wing, eagle:wing, angel:wing, cat:paw, lion:wings, fly:wing
	beer:alcohol	wine:alcohol, cider:alcohol, soda:sugar, beer:liquor, beer:malt, lager:alcoho
	oxygen:atmosphere	helium:atmosphere, hydrogen:atmosphere, nitrogen:atmosphere, methane:atmosphere, carbon:atmosphere
Hypernymy	chihuahua:dog	dachshund:dog, poodle:dog, terrier:dog, chinchilla:rodent, macaque:monkey, dalmatian:dog
	pelican:bird	toucan:bird, puffin:bird, egret:bird, peacock:bird, grouse:bird, pigeon:bird
	tennis:sport	hockey:sport, soccer:sport, volleyball:sport, cricket:sport, golf:sport, football:sport
Morphology	dog:dogs	cat:cats, horse:horses, pig:pigs, rat:rats, wolf:wolves, monkey:monkeys
	tall:tallest	strong:strongest, short:shortest, smart:smartest, weak:weakest, big:biggest, small:smallest
	spy:espionage	pirate:piracy, robber:robbery, lobbyist:lobbying, scout:scouting, terrorist:terrorism, witch:witchcraft

Examples

Category	Target	Nearest Neighbors RelBERT
Commonsense	barista:coffee	baker:bread, brewer:beer, bartender:cocktail, winemaker:wine, bartender:drink, baker:cake
	restaurant:waitress	restaurant:waiter, diner:waitress, bar:bartender, hospital:nurse, courthouse:clerk, office:clerk
	car:garage	car:pit, plane:hangar, auto:garage, baby:crib, yacht:harbour, aircraft:hangar
	ice:melt	snow:melt, glacier:melt, ice:drift, crust:melt, polar ice:melt, ice:thaw
	dolphin:swim	squid:swim, salmon:swim, shark:swim, fish:swim, horse:run, frog:leap
	flower:fragrant	orchid:fragrant, cluster:fragrant, jewel:precious, jewel:valuable, soil:permeable, vegetation:abundant
	coconut:milk	coconut:oil, goat:milk, grape:juice, palm:oil, olive:oil, camel:milk
	bag:plastic	bottle:plastic, bag:leather, container:plastic, box:plastic, jug:glass, bottle:glass
	duck:duckling	chicken:chick, pig:piglet, cat:kitten, ox:calf, butterfly:larvae, bear:cub
Gender	man:woman	men:women, male:female, father:mother, boy:girl, hero:heroine, king:queen
Antonymy	cooked:raw	raw:cooked, regulated:unregulated, sober:drunk, loaded:unloaded, armed:unarmed, published:unpublished
	normal:abnormal	ordinary:unusual, usual:unusual, acceptable:unacceptable, stable:unstable, rational:irrational, legal:illegal
Meronymy	helicopter:rotor	helicopter:rotor blades, helicopter:wing, bicycle:wheel, motorcycle:wheel, airplane:engine, plane:engine
	bat:wing	butterfly:wing, eagle:wing, angel:wing, cat:paw, lion:wings, fly:wing
	beer:alcohol	wine:alcohol, cider:alcohol, soda:sugar, beer:liquor, beer:malt, lager:alcoho
	oxygen:atmosphere	helium:atmosphere, hydrogen:atmosphere, nitrogen:atmosphere, methane:atmosphere, carbon:atmosphere
Hypernymy	chihuahua:dog	dachshund:dog, poodle:dog, terrier:dog, chinchilla:rodent, macaque:monkey, dalmatian:dog
	pelican:bird	toucan:bird, puffin:bird, egret:bird, peacock:bird, grouse:bird, pigeon:bird
	tennis:sport	hockey:sport, soccer:sport, volleyball:sport, cricket:sport, golf:sport, football:sport
Morphology	dog:dogs	cat:cats, horse:horses, pig:pigs, rat:rats, wolf:wolves, monkey:monkeys
	tall:tallest	strong:strongest, short:shortest, smart:smartest, weak:weakest, big:biggest, small:smallest
	spy:espionage	pirate:piracy, robber:robbery, lobbyist:lobbying, scout:scouting, terrorist:terrorism, witch:witchcraft

Examples

Category	Target	Nearest Neighbors RelBERT
Commonsense	barista:coffee	baker:bread, brewer:beer, bartender:cocktail, winemaker:wine, bartender:drink, baker:cake
	restaurant:waitress	restaurant:waiter, diner:waitress, bar:bartender, hospital:nurse, courthouse:clerk, office:clerk
	car:garage	car:pit, plane:hangar, auto:garage, baby:crib, yacht:harbour, aircraft:hangar
	ice:melt	snow:melt, glacier:melt, ice:drift, crust:melt, polar ice:melt, ice:thaw
	dolphin:swim	squid:swim, salmon:swim, shark:swim, fish:swim, horse:run, frog:leap
	flower:fragrant	orchid:fragrant, cluster:fragrant, jewel:precious, jewel:valuable, soil:permeable, vegetation:abundant
	coconut:milk	coconut:oil, goat:milk, grape:juice, palm:oil, olive:oil, camel:milk
	bag:plastic	bottle:plastic, bag:leather, container:plastic, box:plastic, jug:glass, bottle:glass
	duck:duckling	chicken:chick, pig:piglet, cat:kitten, ox:calf, butterfly:larvae, bear:cub
Gender	man:woman	men:women, male:female, father:mother, boy:girl, hero:heroine, king:queen
Antonymy	cooked:raw	raw:cooked, regulated:unregulated, sober:drunk, loaded:unloaded, armed:unarmed, published:unpublished
	normal:abnormal	ordinary:unusual, usual:unusual, acceptable:unacceptable, stable:unstable, rational:irrational, legal:illegal
Meronymy	helicopter:rotor	helicopter:rotor blades, helicopter:wing, bicycle:wheel, motorcycle:wheel, airplane:engine, plane:engine
	bat:wing	butterfly:wing, eagle:wing, angel:wing, cat:paw, lion:wings, fly:wing
	beer:alcohol	wine:alcohol, cider:alcohol, soda:sugar, beer:liquor, beer:malt, lager:alcoho
	oxygen:atmosphere	helium:atmosphere, hydrogen:atmosphere, nitrogen:atmosphere, methane:atmosphere, carbon:atmosphere
Hypernymy	chihuahua:dog	dachshund:dog, poodle:dog, terrier:dog, chinchilla:rodent, macaque:monkey, dalmatian:dog
	pelican:bird	toucan:bird, puffin:bird, egret:bird, peacock:bird, grouse:bird, pigeon:bird
	tennis:sport	hockey:sport, soccer:sport, volleyball:sport, cricket:sport, golf:sport, football:sport
Morphology	dog:dogs	cat:cats, horse:horses, pig:pigs, rat:rats, wolf:wolves, monkey:monkeys
	tall:tallest	strong:strongest, short:shortest, smart:smartest, weak:weakest, big:biggest, small:smallest
	spy:espionage	pirate:piracy, robber:robbery, lobbyist:lobbying, scout:scouting, terrorist:terrorism, witch:witchcraft