

How **AI** UNDERSTANDS LANGUAGE

 Follow @cocoweixu

Wei Xu

Department of Computer Science and Engineering



THE OHIO STATE UNIVERSITY

THE TURING TEST

VOL. LIX. No. 236.]

[October, 1950]

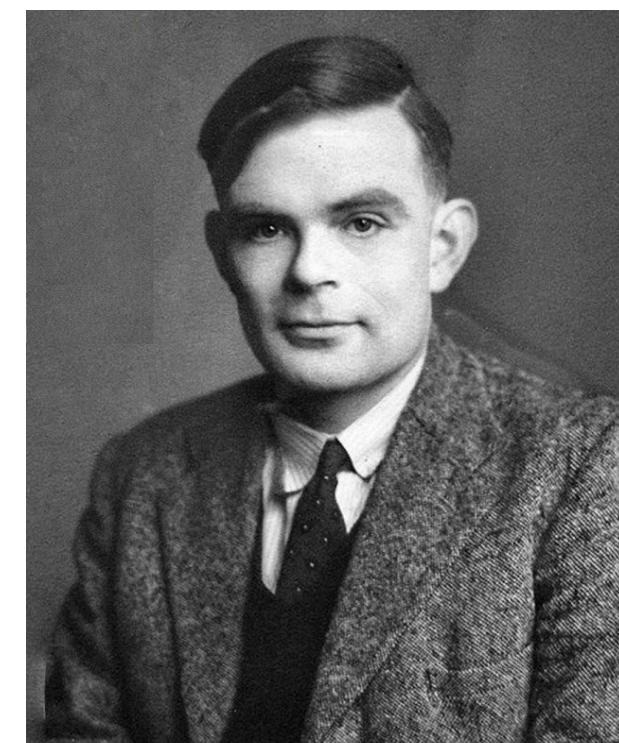
MIND
A QUARTERLY REVIEW
OF
PSYCHOLOGY AND PHILOSOPHY

I.—COMPUTING MACHINERY AND
INTELLIGENCE

By A. M. TURING

1. *The Imitation Game.*

I PROPOSE to consider the question, ‘Can machines think ?’ This should begin with definitions of the meaning of the terms ‘machine’ and ‘think’. The definitions might be framed so as to reflect so far as possible the normal use of the words, but this attitude is dangerous. If the meaning of the words ‘machine’ and ‘think’ are to be found by examining how they are commonly used it is difficult to escape the conclusion that the meaning and the answer to the question, ‘Can machines think ?’ is to be sought in a statistical survey such as a Gallup poll. But this is absurd. Instead of attempting such a definition I shall replace the question by another, which is closely related to it and is expressed in relatively unambiguous words.



Alan Turing (1912-1954) mathematician, world-class Marathon runner, forced to chemical castration for homosexuality conviction.

THE TURING TEST

VOL. LIX. No. 236.]

[October, 1950]

‘Can machines think?’

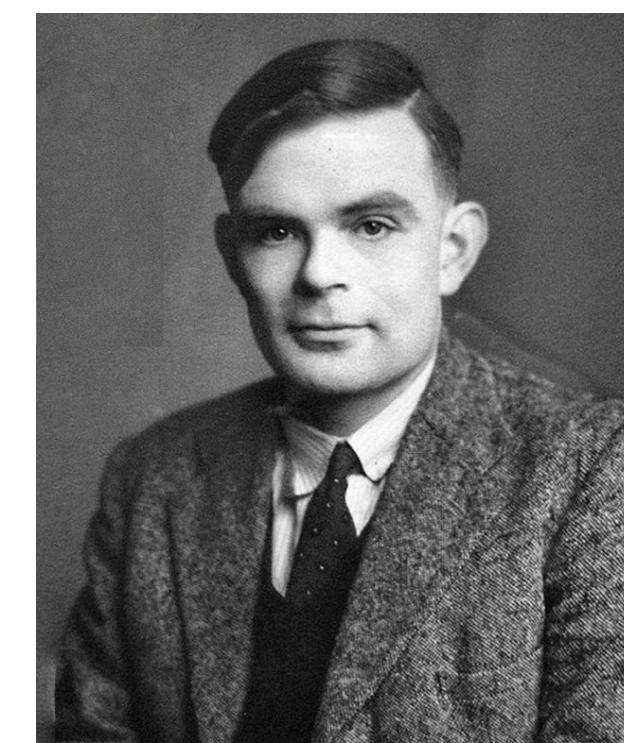
MIND
A QUARTERLY REVIEW
OF
PSYCHOLOGY AND PHILOSOPHY

I.—COMPUTING MACHINERY AND
INTELLIGENCE

By A. M. TURING

1. *The Imitation Game.*

I PROPOSE to consider the question, ‘Can machines think?’ This should begin with definitions of the meaning of the terms ‘machine’ and ‘think’. The definitions might be framed so as to reflect so far as possible the normal use of the words, but this attitude is dangerous. If the meaning of the words ‘machine’ and ‘think’ are to be found by examining how they are commonly used it is difficult to escape the conclusion that the meaning and the answer to the question, ‘Can machines think?’ is to be sought in a statistical survey such as a Gallup poll. But this is absurd. Instead of attempting such a definition I shall replace the question by another, which is closely related to it and is expressed in relatively unambiguous words.



Alan Turing (1912-1954) mathematician, world-class Marathon runner, forced to chemical castration for homosexuality conviction.

THE TURING TEST

VOL. LIX. No. 236.]

[October, 1950]

‘Can machines think?’

M I N D
A QUARTERLY REVIEW
OF
PSYCHOLOGY AND PHILOSOPHY

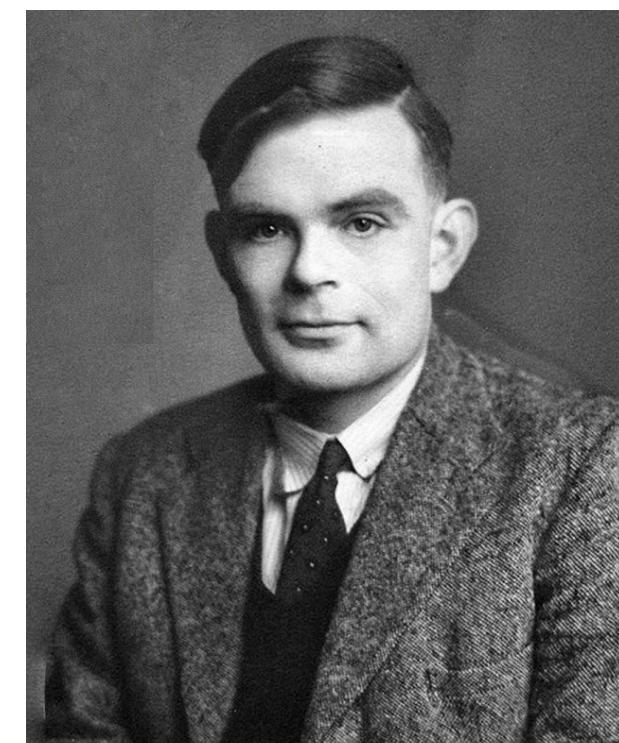
I.—COMPUTING MACHINERY AND
INTELLIGENCE

By A. M. TURING

1. *The Imitation Game.*

I PROPOSE to consider the question, ‘Can machines think?’ This should begin with definitions of the meaning of the terms ‘machine’ and ‘think’. The definitions might be framed so as to reflect so far as possible the normal use of the words, but this attitude is dangerous. If the meaning of the words ‘machine’ and ‘think’ are to be found by examining how they are commonly used it is difficult to escape the conclusion that the meaning and the answer to the question, ‘Can machines think?’ is to be sought in a statistical survey such as a Gallup poll. But this is absurd. Instead of attempting such a definition I shall replace the question by another, which is closely related to it and is expressed in relatively unambiguous words.

I shall replace the question by another,



Alan Turing (1912-1954) mathematician, world-class Marathon runner, forced to chemical castration for homosexuality conviction.

THE TURING TEST

VOL. LIX. No. 236.]

[October, 1950]

‘Can machines think?’

M I N D
A QUARTERLY REVIEW
OF
PSYCHOLOGY AND PHILOSOPHY

I.—COMPUTING MACHINERY AND
INTELLIGENCE

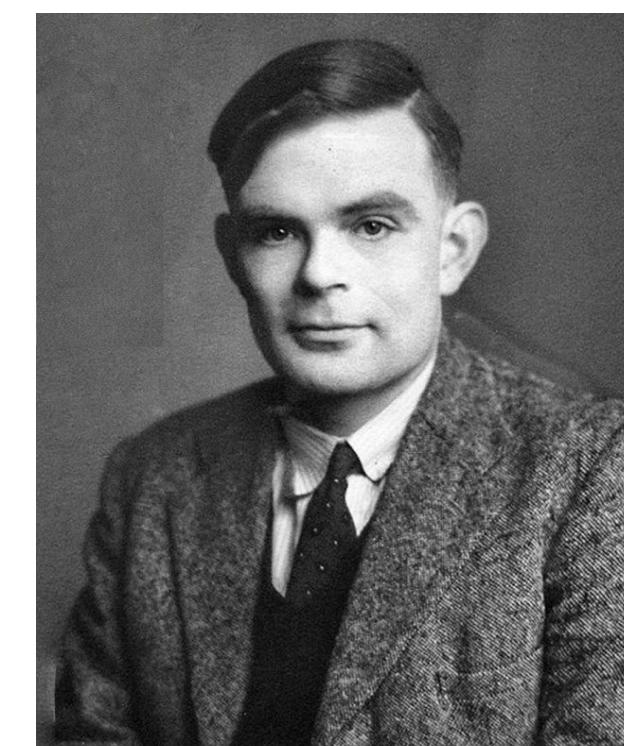
By A. M. TURING

1. *The Imitation Game.*

I PROPOSE to consider the question, ‘Can machines think?’ This should begin with definitions of the meaning of the terms ‘machine’ and ‘think’. The definitions might be framed so as to reflect so far as possible the normal use of the words, but this attitude is dangerous. If the meaning of the words ‘machine’ and ‘think’ are to be found by examining how they are commonly used it is difficult to escape the conclusion that the meaning and the answer to the question, ‘Can machines think?’ is to be sought in a statistical survey such as a Gallup poll. But this is absurd. Instead of attempting such a definition I shall replace the question by another, which is closely related to it and is expressed in relatively unambiguous words.

I shall replace the question by another,

Can machine conduct natural language
conversations like a human?



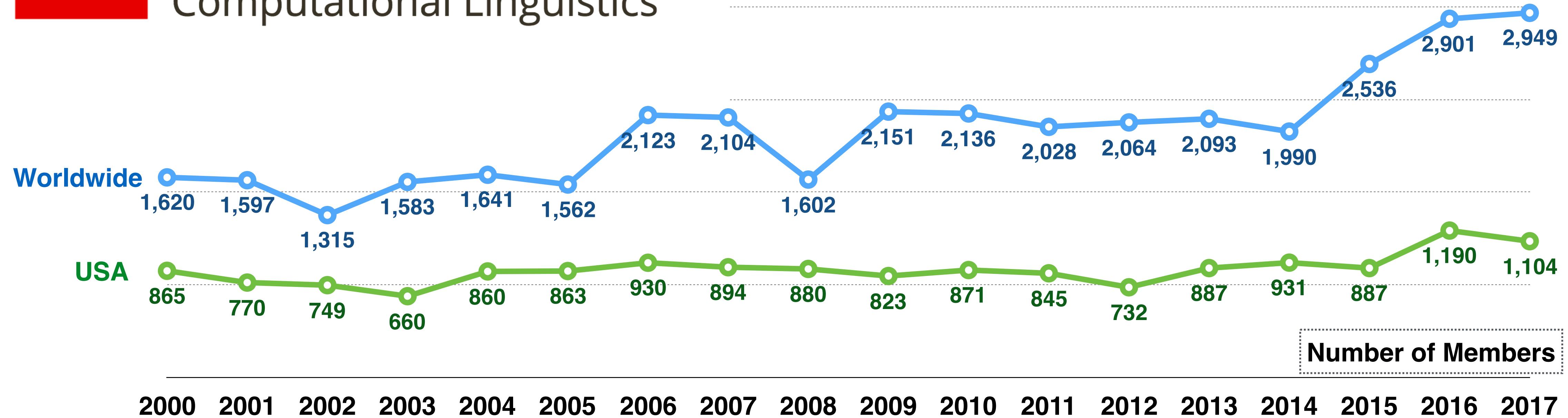
Alan Turing (1912-1954) mathematician, world-class Marathon runner, forced to chemical castration for homosexuality conviction.

NATURAL LANGUAGE PROCESSING (NLP)

The goal is to enable computers to understand and generate human language.



Association for
Computational Linguistics



Number of Members

Human language is **ambiguous**, creative, infinite, and ever evolving.



Students Cook & Serve Grandparents

On Thursday, September 9, Gorman School hosted the first annual Grandparent's Day. All Grandparents were invited to a school wide pancake breakfast. Upper grade students served as excellent chefs, as well as taking responsibility for serving the food and the clean up after-

THE HOLY GRAIL OF AI / NLP

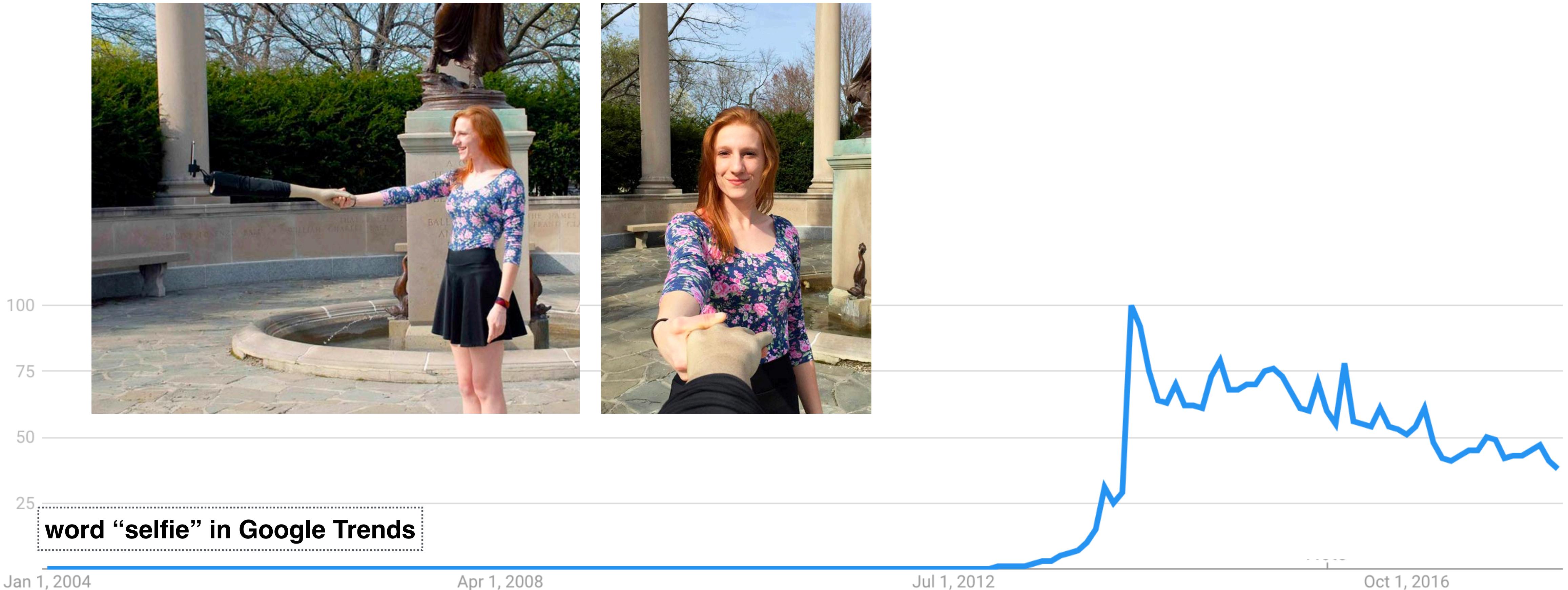
Human language is ambiguous, **creative, infinite**, and ever evolving.

Jeremy Corbyn is a closet Brexiteer. = Mr Corbyn is actually a secret supporter of Brexit.



THE HOLY GRAIL OF AI / NLP

Human language is ambiguous, creative, infinite, and **ever evolving**.



THE HOLY GRAIL OF AI / NLP

Human language is ambiguous, creative, infinite, and ever evolving.

Solution: learning large-scale paraphrases

THE HOLY GRAIL OF AI / NLP

Human language is ambiguous, creative, infinite, and ever evolving.

Solution: learning large-scale paraphrases

selfie

word

photo

Human language is ambiguous, creative, infinite, and ever evolving.

Solution: learning large-scale paraphrases

selfie **word** photo

gets the boot from **phrase** has been sacked by

Human language is ambiguous, creative, infinite, and ever evolving.

Solution: learning large-scale paraphrases

selfie **word** photo

gets the boot from **phrase** has been sacked by

Jeremy Corbyn is a closet Brexiteer. **sentence** Mr Corbyn is actually a secret supporter of Brexit.



Question Answering

Who is the CEO **stepping down** from Boeing?



Question Answering

Who is the CEO **stepping down** from Boeing?



match



Question Answering

Who is the CEO **stepping down** from Boeing?

match

… the forced **resignation** of the CEO
of Boeing, Harry Stonecipher, for …



Question Answering

Who is the CEO **stepping down** from Boeing?

match

… the forced **resignation** of the CEO
of Boeing, Harry Stonecipher, for …



Question Answering

Who is the CEO **stepping down** from Boeing?

match

… the forced **resignation** of the CEO
of Boeing, Harry Stonecipher, for …

… after Boeing Co. Chief Executive
Harry Stonecipher was **ousted** from …

NATURAL LANGUAGE GENERATION

Text Simplification

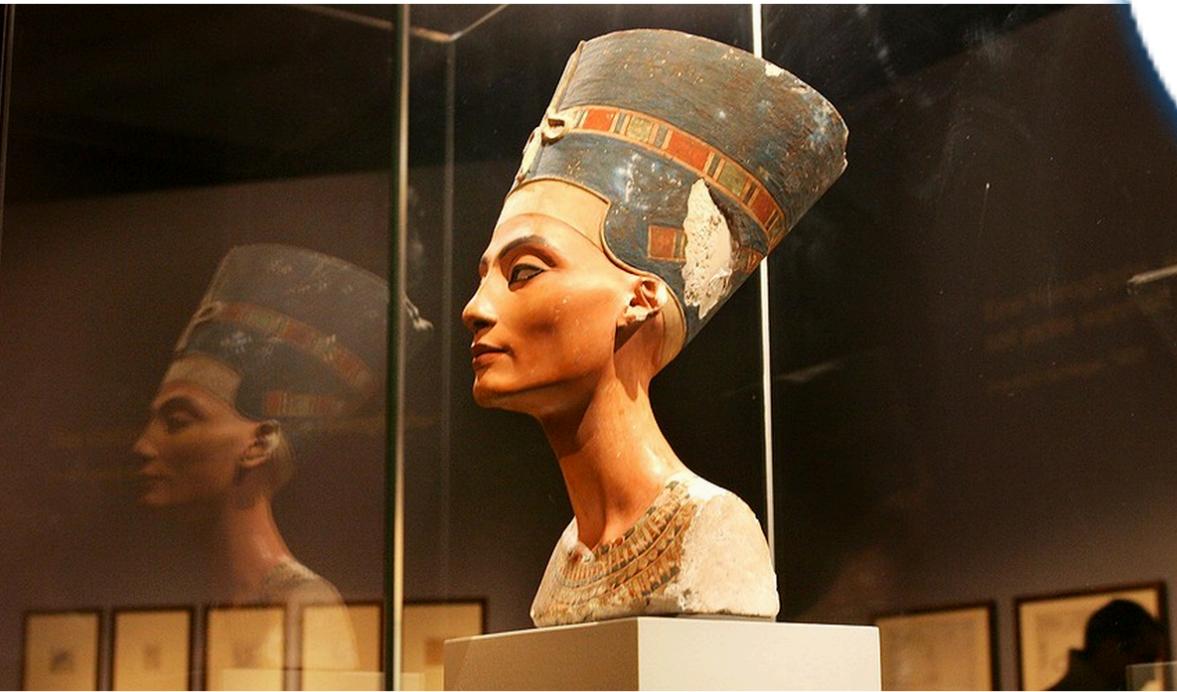
NEWSela

WAR & PEACE SCIENCE KIDS MONEY HEALTH

SCIENCE 1738

Archaeologist may have found remains of ancient Egyptian Queen Nefertiti

By Robert Gebelhoff, Washington Post.
08.17.15



The 3,330-year-old bust of Nefertiti sits in an exhibition in the Kulturforum in Berlin, Germany, March 1, 2005.
Photo: AP/Herbert Knosowski

Nefertiti — she's an ancient Egyptian queen and the source of a fantastic mystery regarding the iconic remnants of long-lost royalty.

For decades, archaeologists have speculated on the location of the queen's remains, the last royal mummy missing from the dynasty of the famous King Tutankhamun, better known as King Tut. But now, an archaeologist claims that he has found her

MAX
1140L

960L
720L
420L

WRITE
 QUIZ

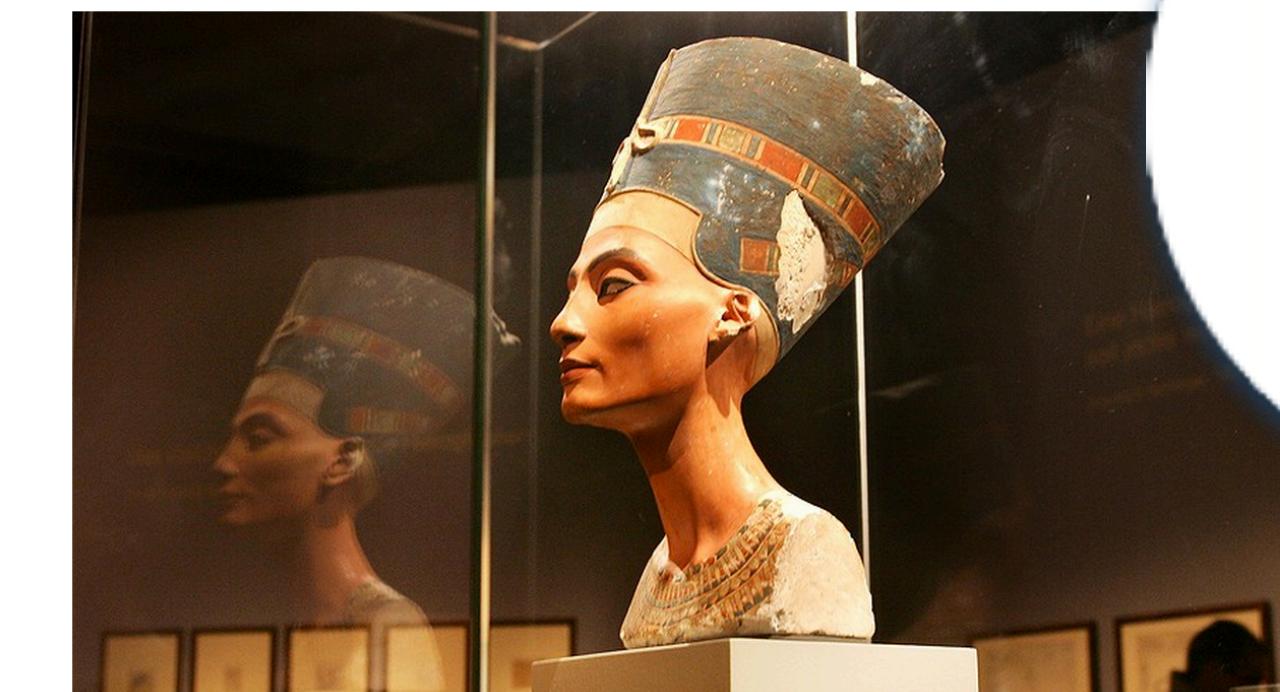
NEWSela

WAR & PEACE SCIENCE KIDS MONEY LAW HEALTH

SCIENCE 1738

Mystery of ancient Egypt solved? Tomb of queen may be hidden near King Tut

By Washington Post, adapted by Newsela staff
08.17.15



The 3,330-year-old bust of Nefertiti sits in an exhibition in the Kulturforum in Berlin, Germany, March 1, 2005.
Photo: AP/Herbert Knosowski

The ancient Egyptian Queen Nefertiti has long been at the center of a mystery. For years, archaeologists have wondered where her tomb might be hidden. Nefertiti belonged to the family line of the famous King Tutankhamun, better known as King Tut. Indeed, some believe she was Tut's mother. While the other royals in her line are

1140L
960L
720L

420L

WRITE
 QUIZ

(adapting machine translation techniques)

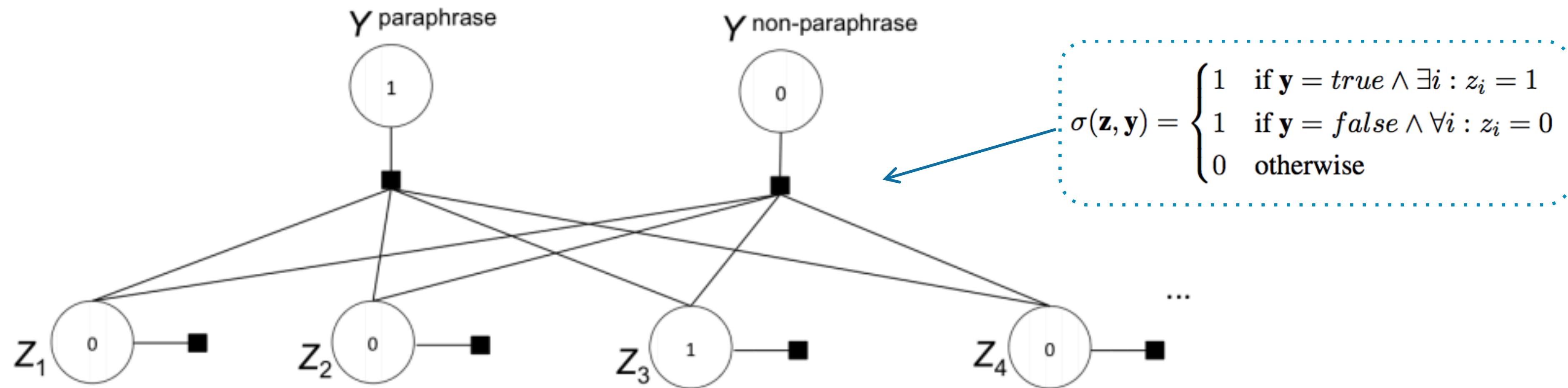
DESIGNING VARIOUS [MACHINE LEARNING] MODELS

DESIGNING VARIOUS [MACHINE LEARNING] MODELS

- **Multi-instance Learning Models** [Xu et al. 2014; Tabassum et al. 2016]

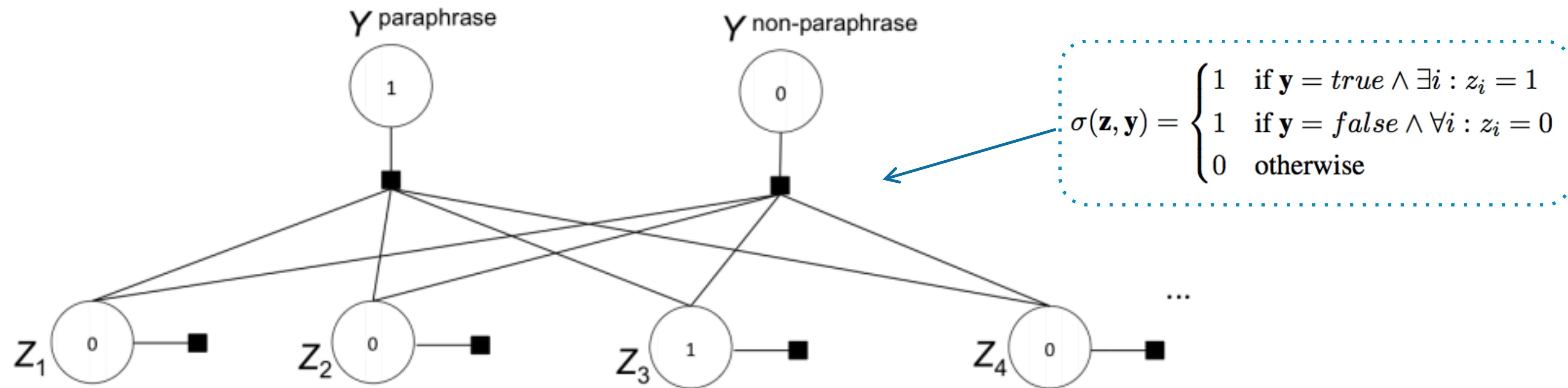
DESIGNING VARIOUS [MACHINE LEARNING] MODELS

- Multi-instance Learning Models [Xu et al. 2014; Tabassum et al. 2016]



DESIGNING VARIOUS [MACHINE LEARNING] MODELS

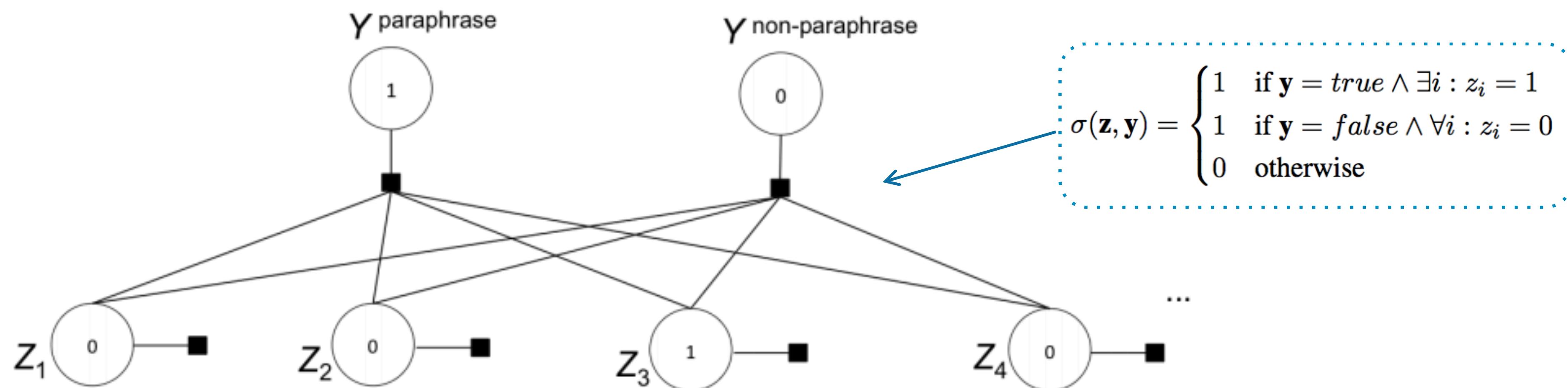
- Multi-instance Learning Models [Xu et al. 2014; Tabassum et al. 2016]



$$P(\mathbf{z}_i, y_i | \mathbf{w}_i; \theta) = \prod_{j=1}^m \exp(\theta \cdot f(z_j, w_j)) \times \sigma(\mathbf{z}_i, y_i)$$

DESIGNING VARIOUS [MACHINE LEARNING] MODELS

- Multi-instance Learning Models [Xu et al. 2014; Tabassum et al. 2016]



$$P(\mathbf{z}_i, y_i | \mathbf{w}_i; \theta) = \prod_{j=1}^m \exp(\theta \cdot f(z_j, w_j)) \times \sigma(\mathbf{z}_i, y_i)$$

$$\begin{aligned} \frac{\partial \log P(\mathbf{y} | \mathbf{w}; \theta)}{\partial \theta} &= \mathbf{E}_{P(\mathbf{z} | \mathbf{w}, \mathbf{y}; \theta)} \left(\sum_i f(\mathbf{z}_i, \mathbf{w}_i) \right) - \mathbf{E}_{P(\mathbf{z}, \mathbf{y} | \mathbf{w}; \theta)} \left(\sum_i f(\mathbf{z}_i, \mathbf{w}_i) \right) \\ &\approx \sum_i f(\mathbf{z}_i^*, \mathbf{w}_i) - \sum_i f(\mathbf{z}'_i, \mathbf{w}_i) \end{aligned}$$

(Viterbi approximation + online learning)

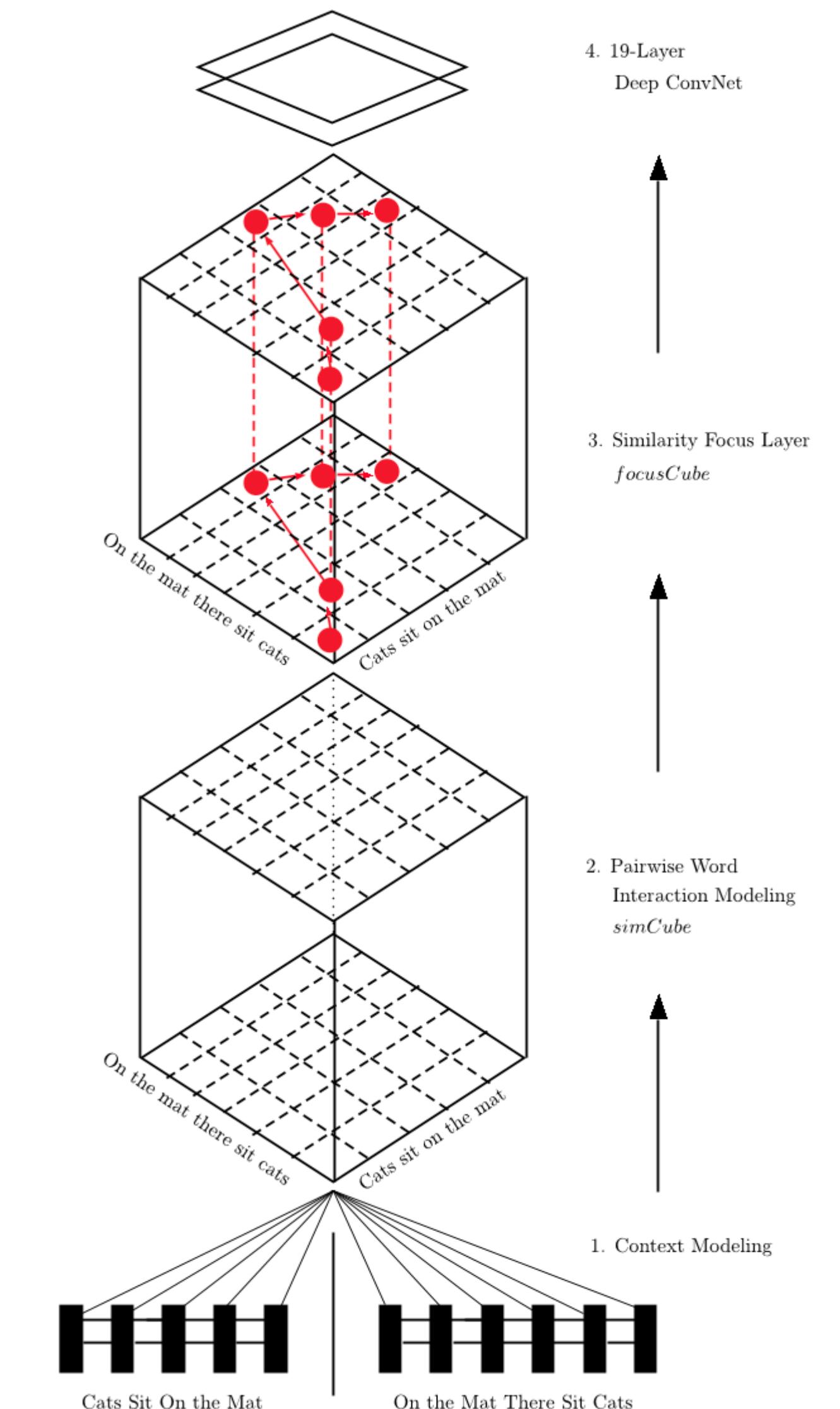
DESIGNING VARIOUS [MACHINE LEARNING] MODELS

DESIGNING VARIOUS [MACHINE LEARNING] MODELS

- **Deep Pairwise Neural Networks** [He et al. 2015; Lan and Xu 2018]

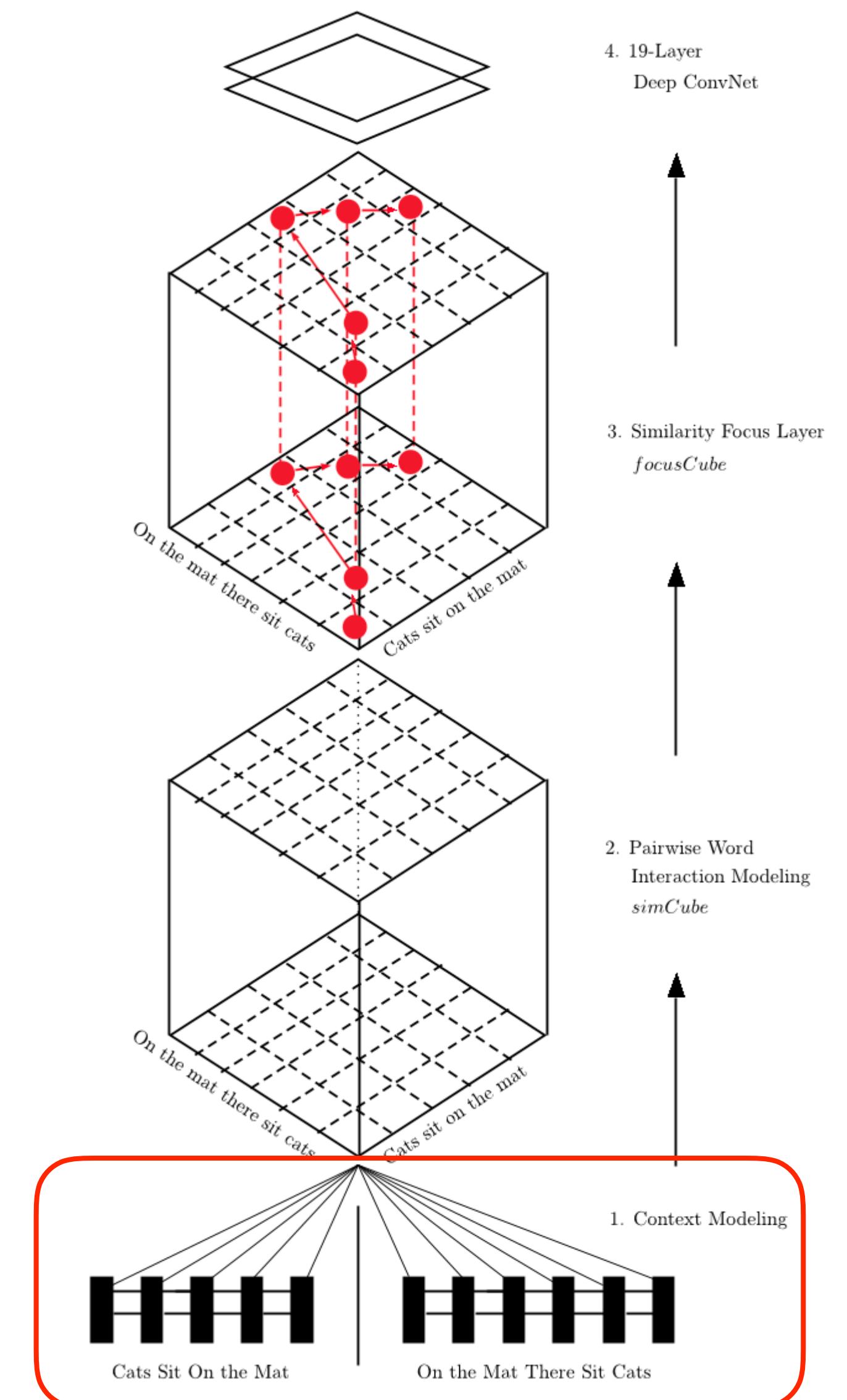
DESIGNING VARIOUS [MACHINE LEARNING] MODELS

- Deep Pairwise Neural Networks [He et al. 2015; Lan and Xu 2018]



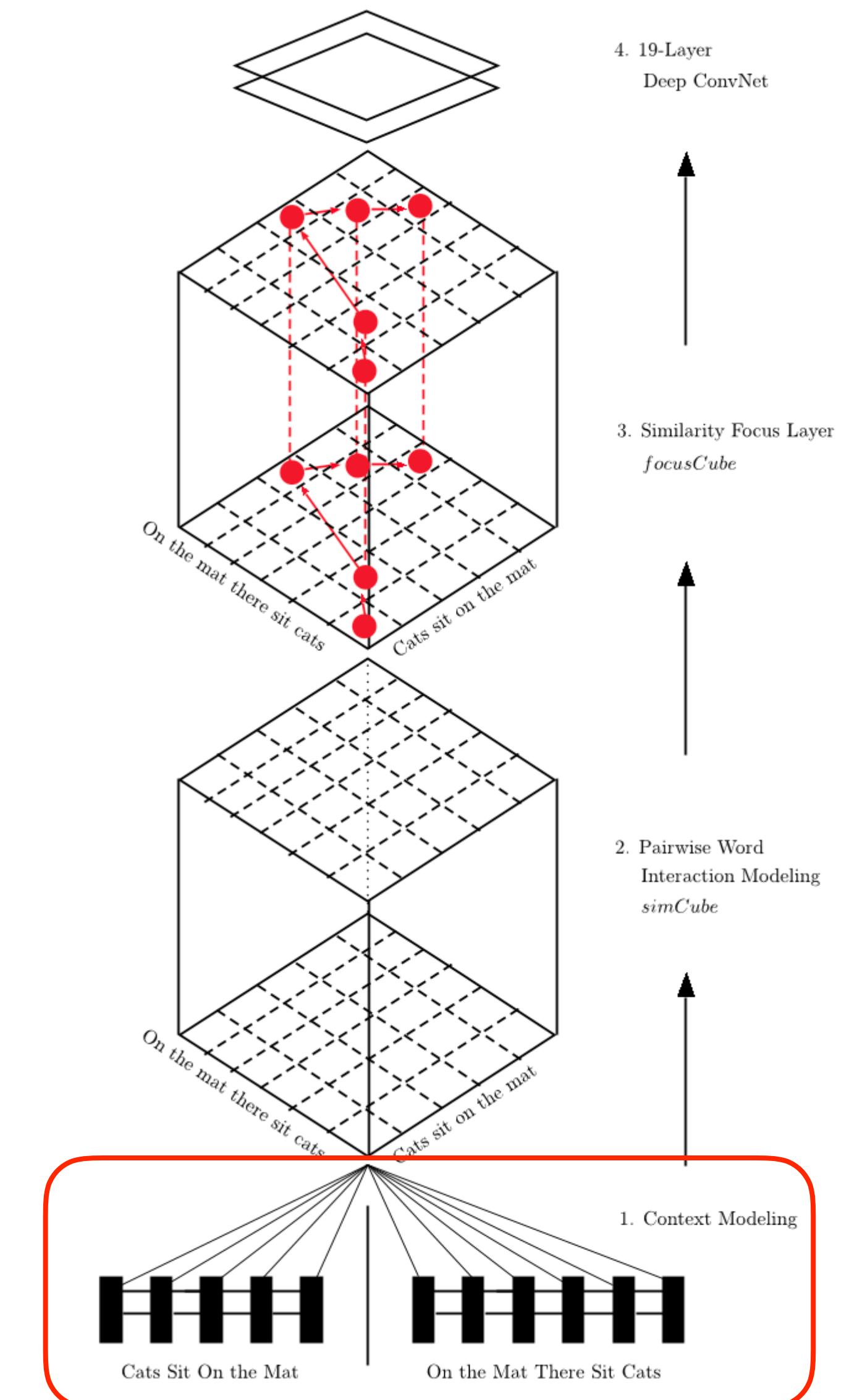
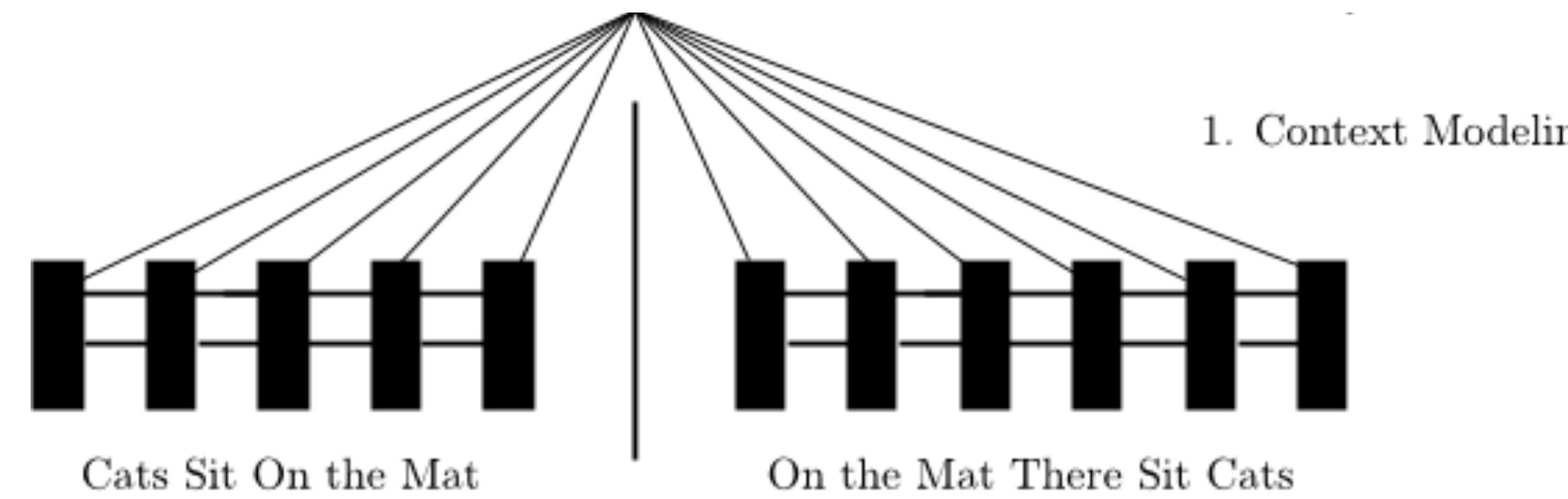
DESIGNING VARIOUS [MACHINE LEARNING] MODELS

- Deep Pairwise Neural Networks [He et al. 2015; Lan and Xu 2018]



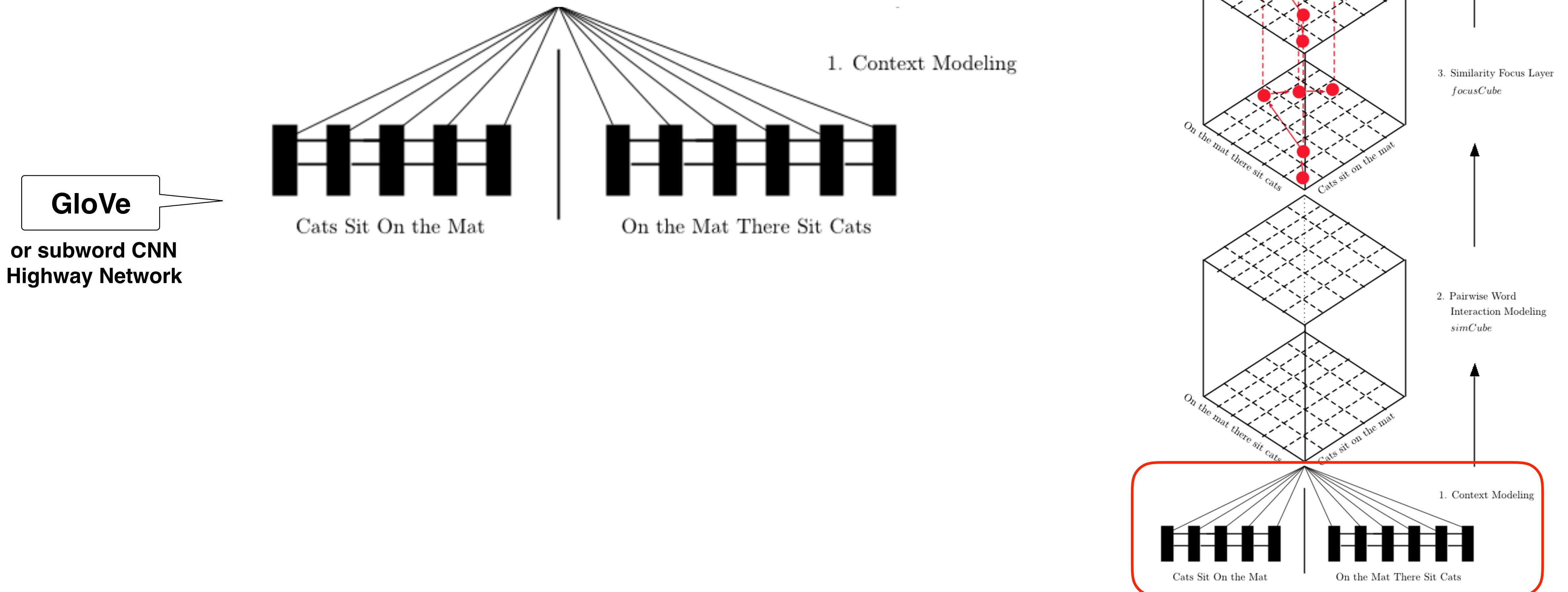
DESIGNING VARIOUS [MACHINE LEARNING] MODELS

- Deep Pairwise Neural Networks [He et al. 2015; Lan and Xu 2018]



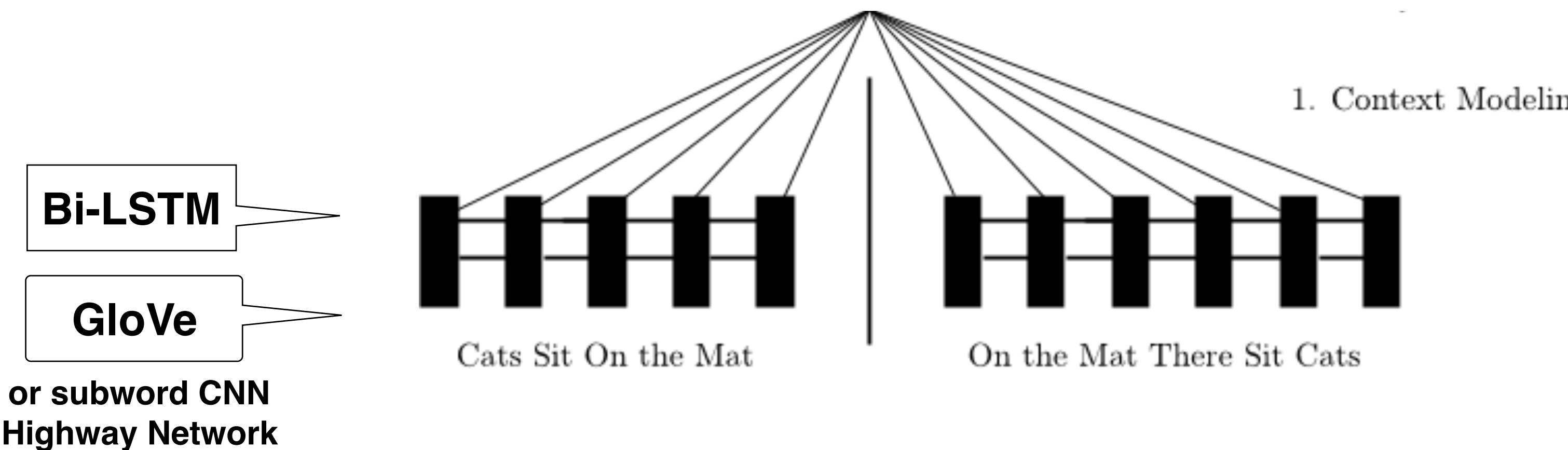
DESIGNING VARIOUS [MACHINE LEARNING] MODELS

- Deep Pairwise Neural Networks [He et al. 2015; Lan and Xu 2018]



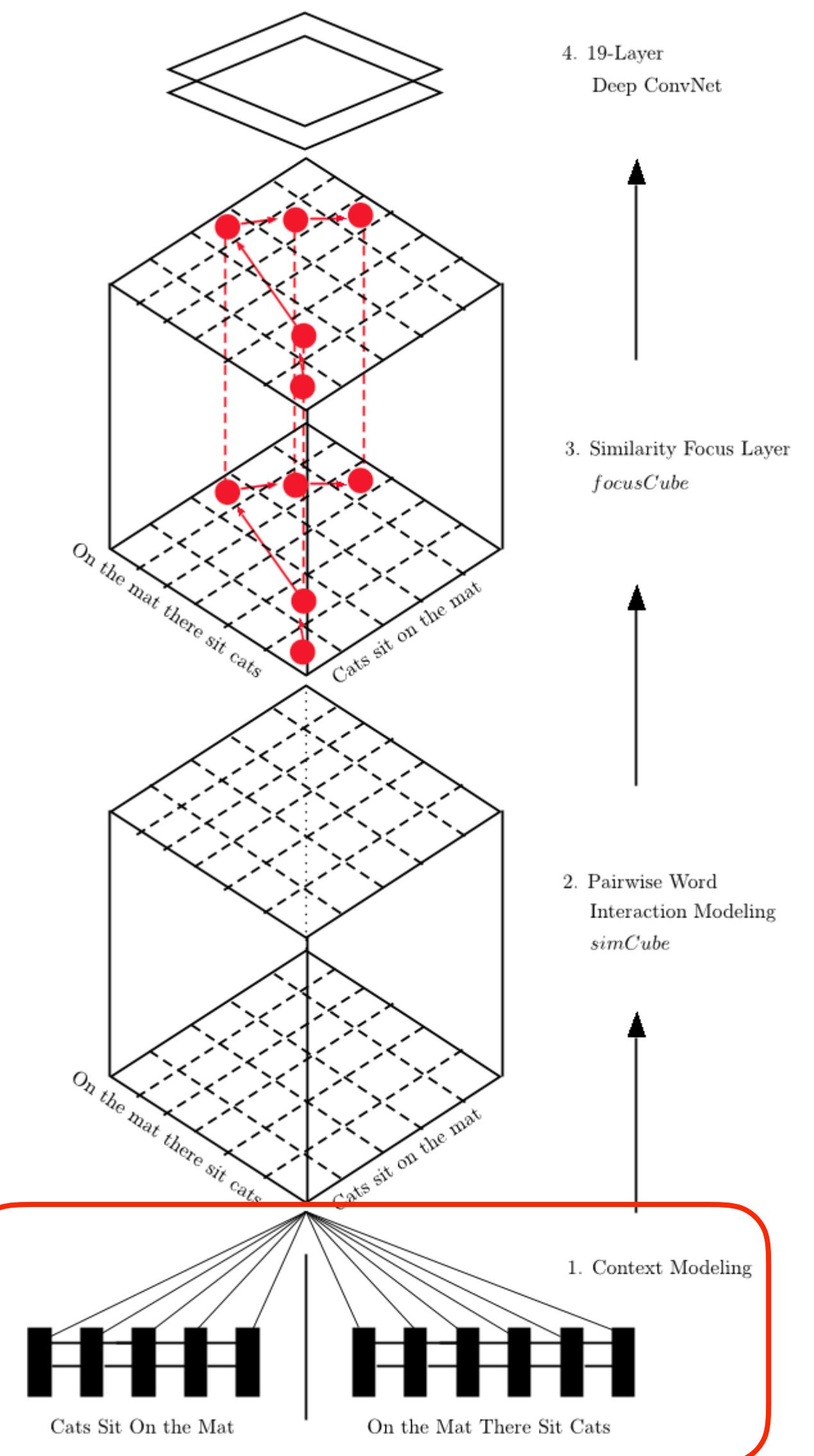
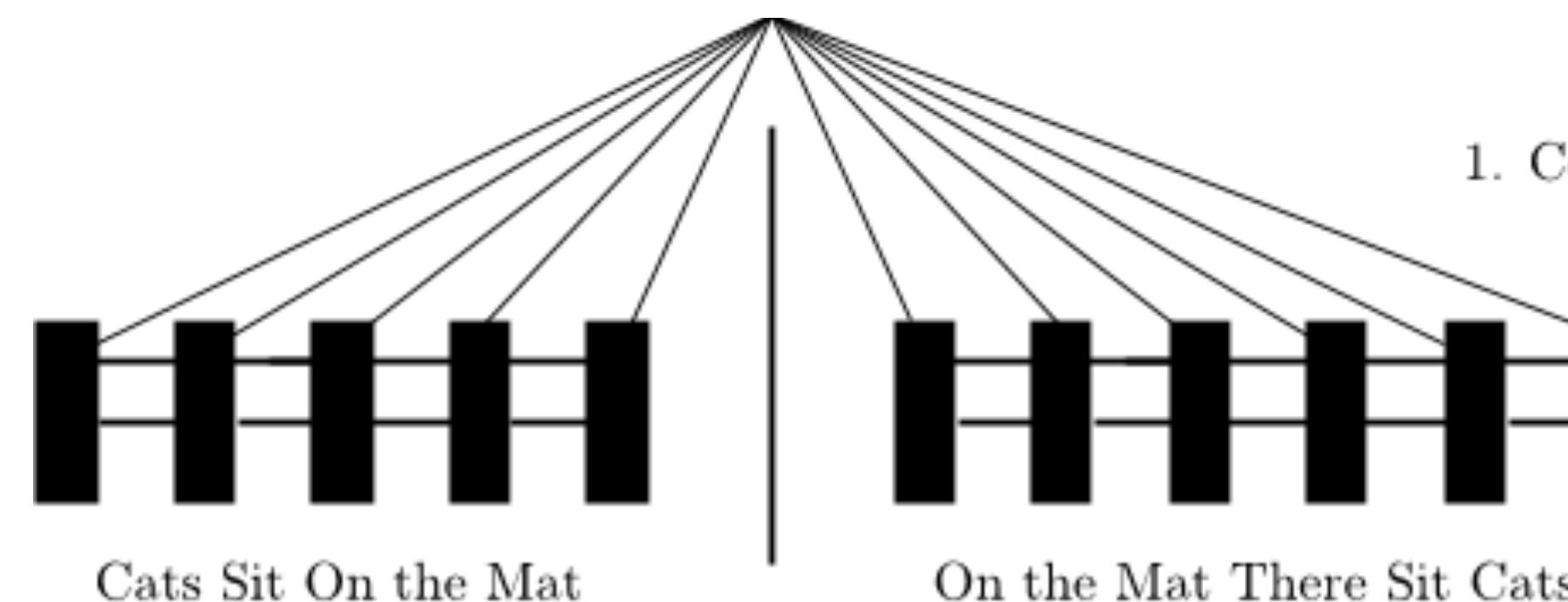
DESIGNING VARIOUS [MACHINE LEARNING] MODELS

- Deep Pairwise Neural Networks [He et al. 2015; Lan and Xu 2018]



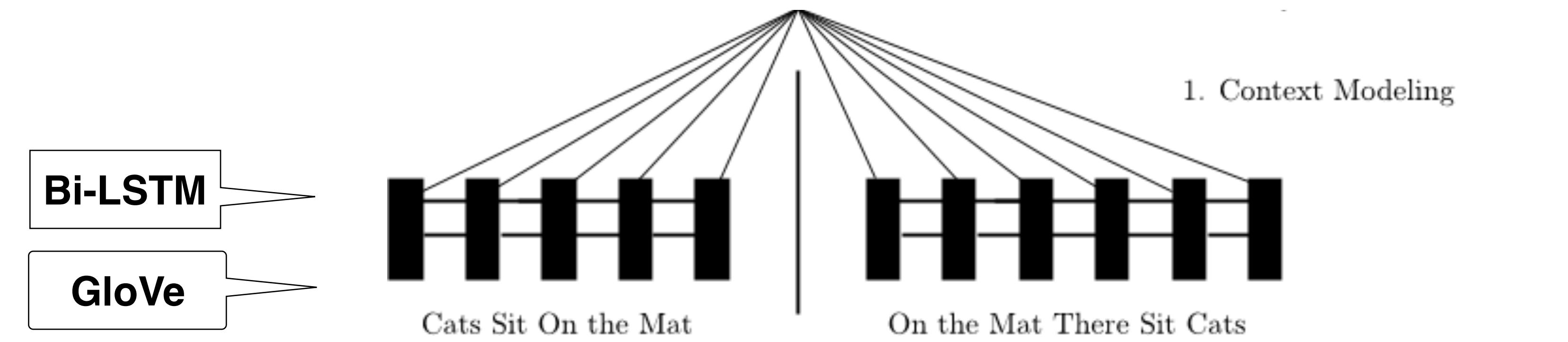
Bi-LSTM
Glove
or subword CNN
Highway Network

1. Context Modeling

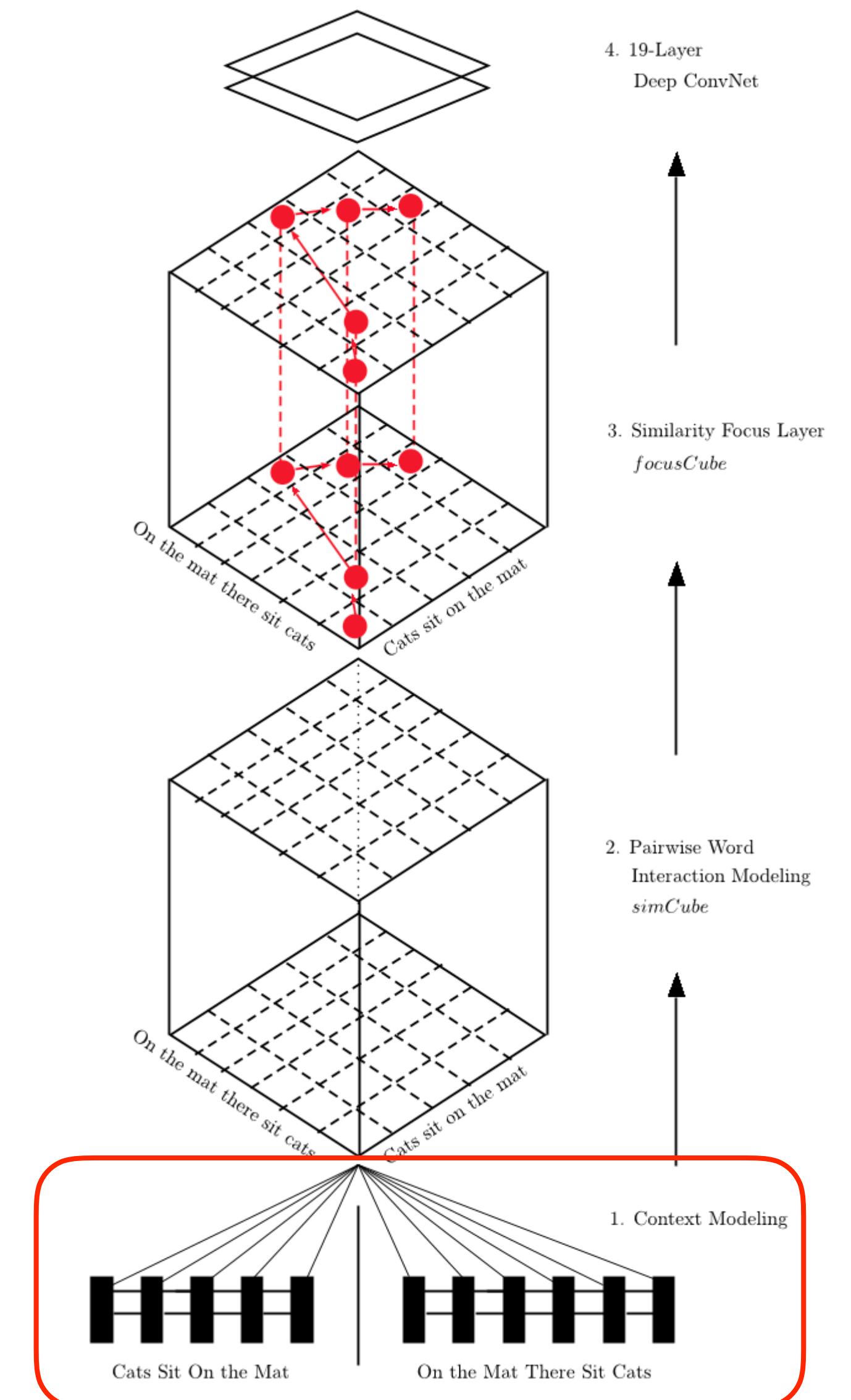


DESIGNING VARIOUS [MACHINE LEARNING] MODELS

- Deep Pairwise Neural Networks [He et al. 2015; Lan and Xu 2018]

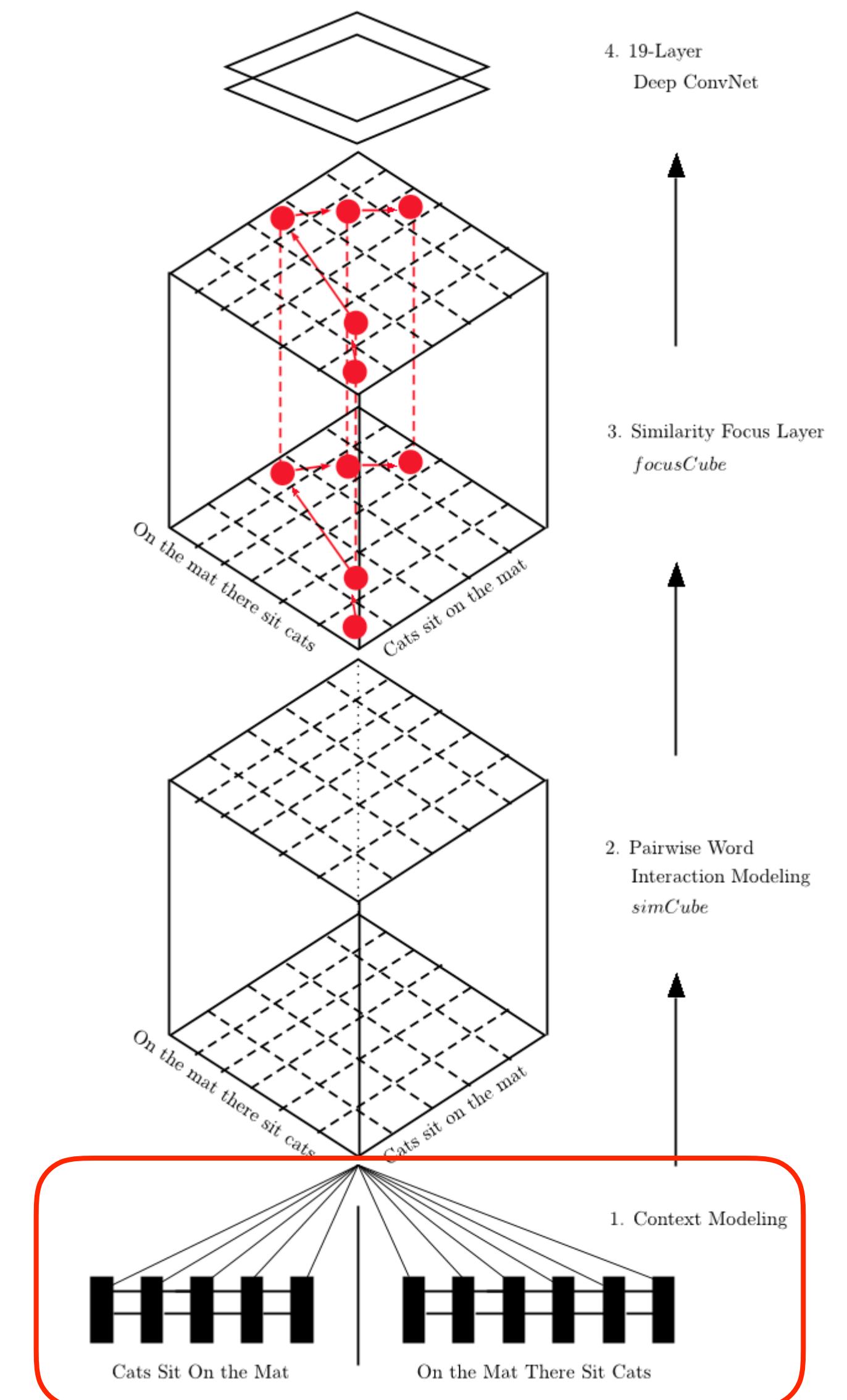


Decompose sentence input into word context to reduce modeling difficulty



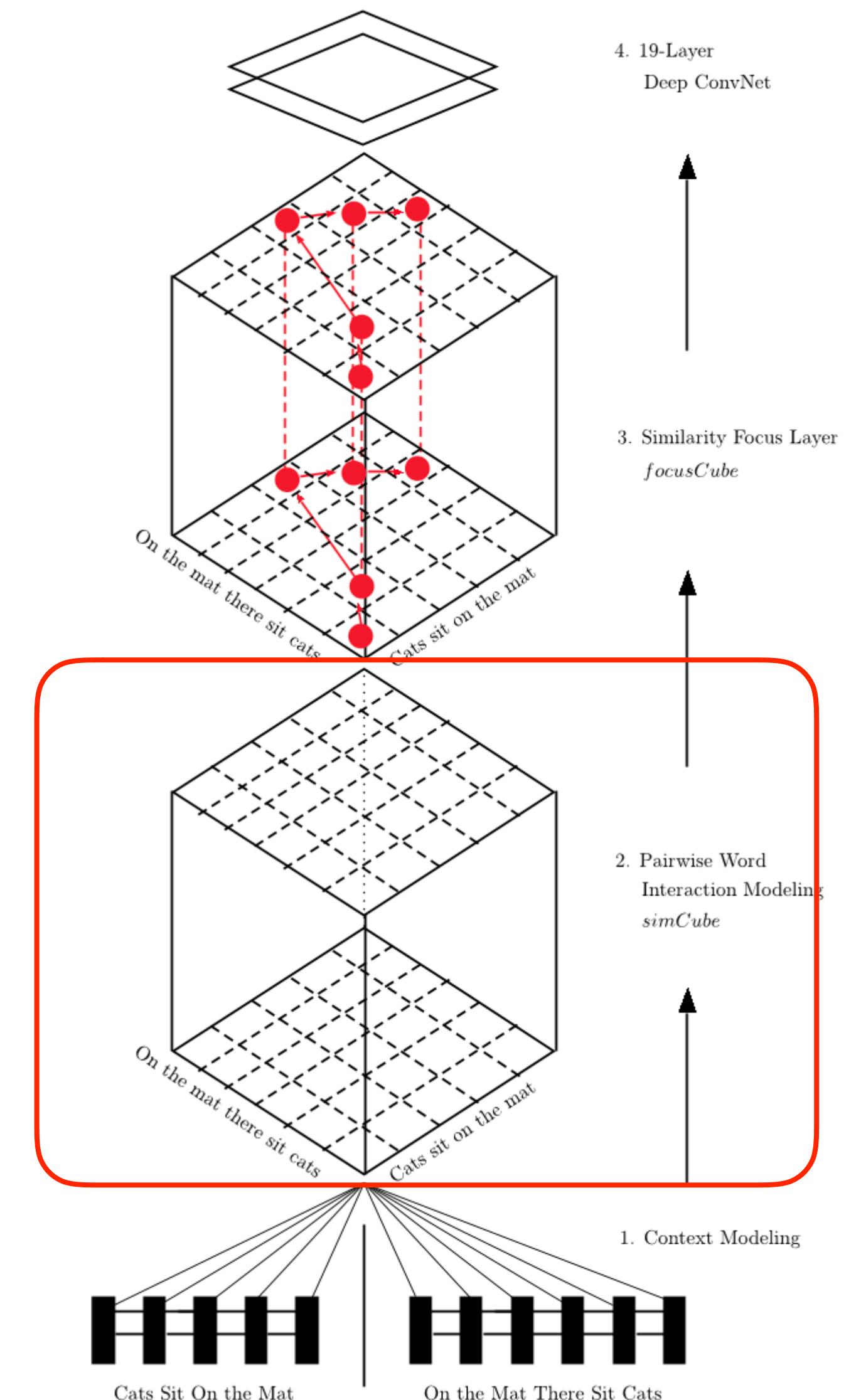
DESIGNING VARIOUS [MACHINE LEARNING] MODELS

- Deep Pairwise Neural Networks [He et al. 2015; Lan and Xu 2018]



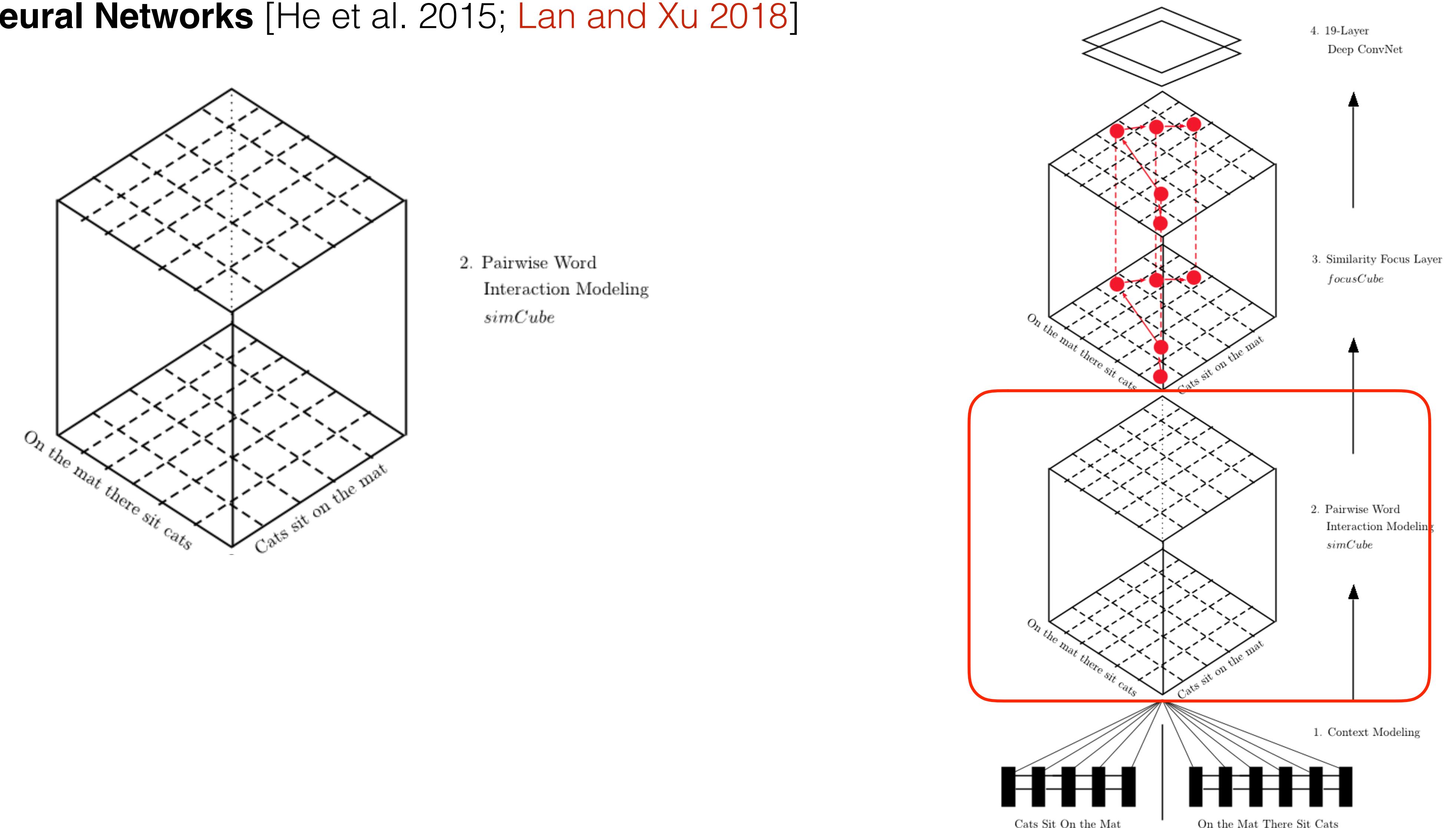
DESIGNING VARIOUS [MACHINE LEARNING] MODELS

- Deep Pairwise Neural Networks [He et al. 2015; Lan and Xu 2018]



DESIGNING VARIOUS [MACHINE LEARNING] MODELS

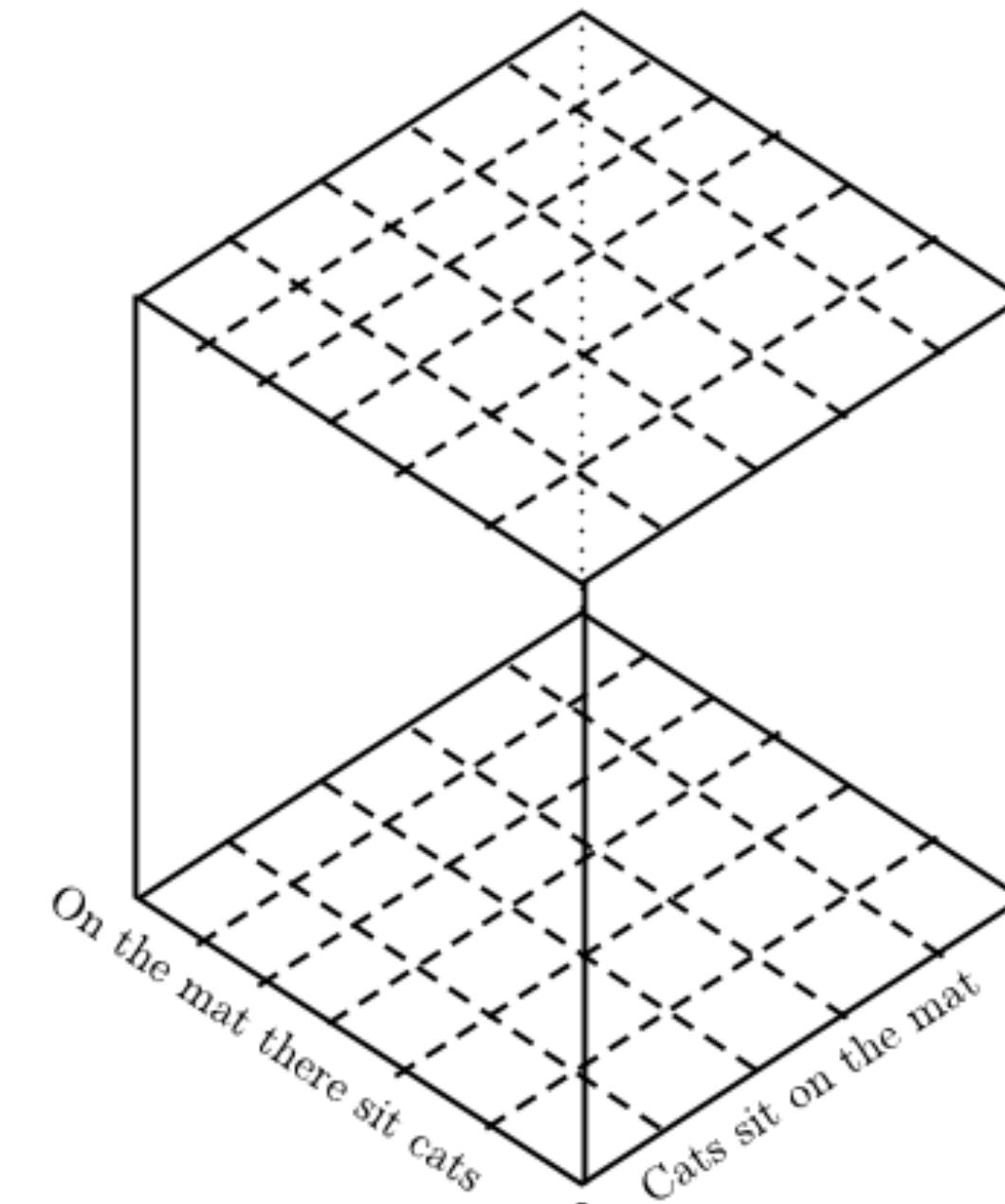
- Deep Pairwise Neural Networks [He et al. 2015; Lan and Xu 2018]



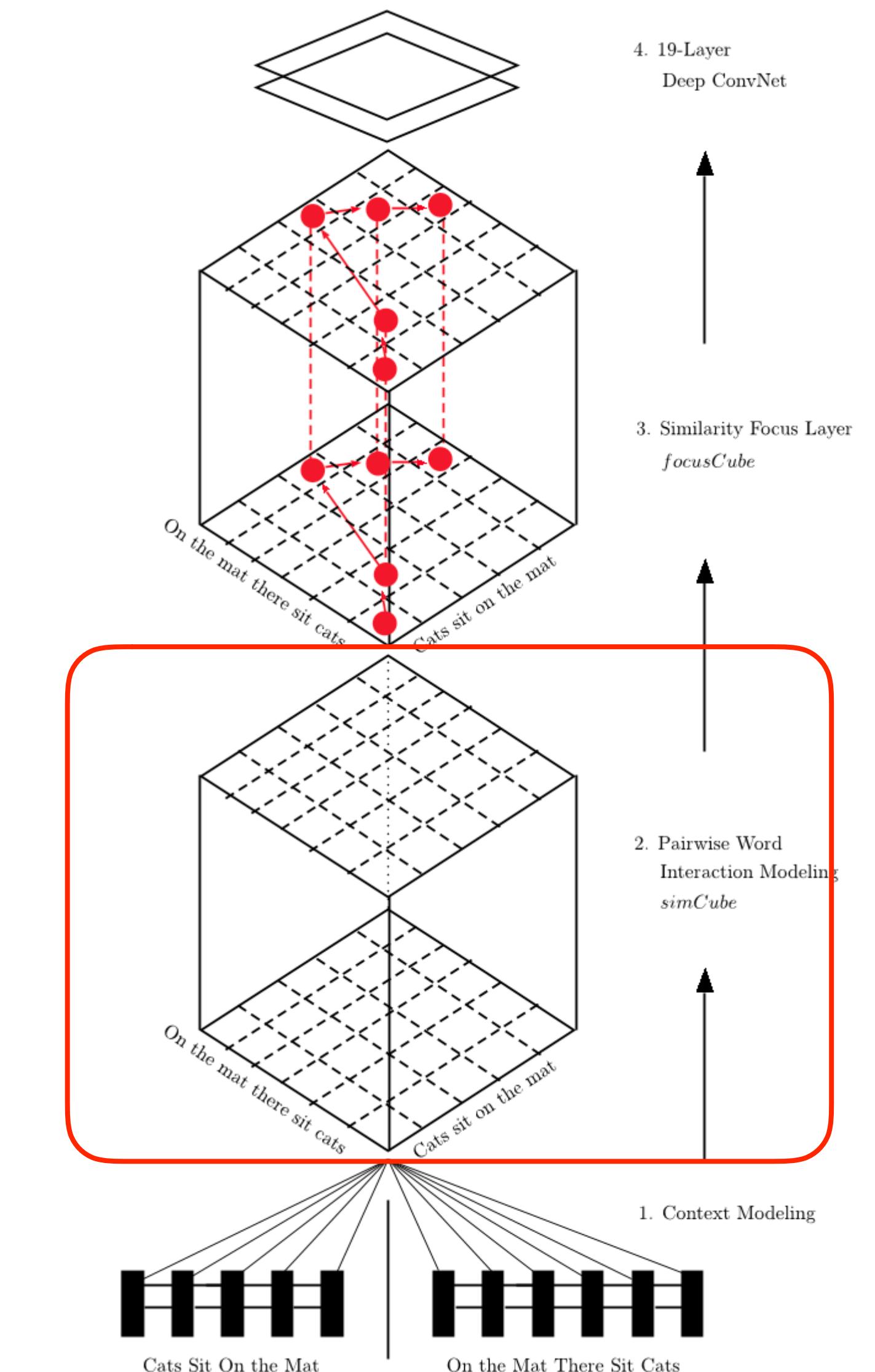
DESIGNING VARIOUS [MACHINE LEARNING] MODELS

- Deep Pairwise Neural Networks [He et al. 2015; Lan and Xu 2018]

$$coU(\vec{h}_1, \vec{h}_2) = \{\cos(\vec{h}_1, \vec{h}_2), L_2 Euclid(\vec{h}_1, \vec{h}_2), DotProduct(\vec{h}_1, \vec{h}_2)\}$$

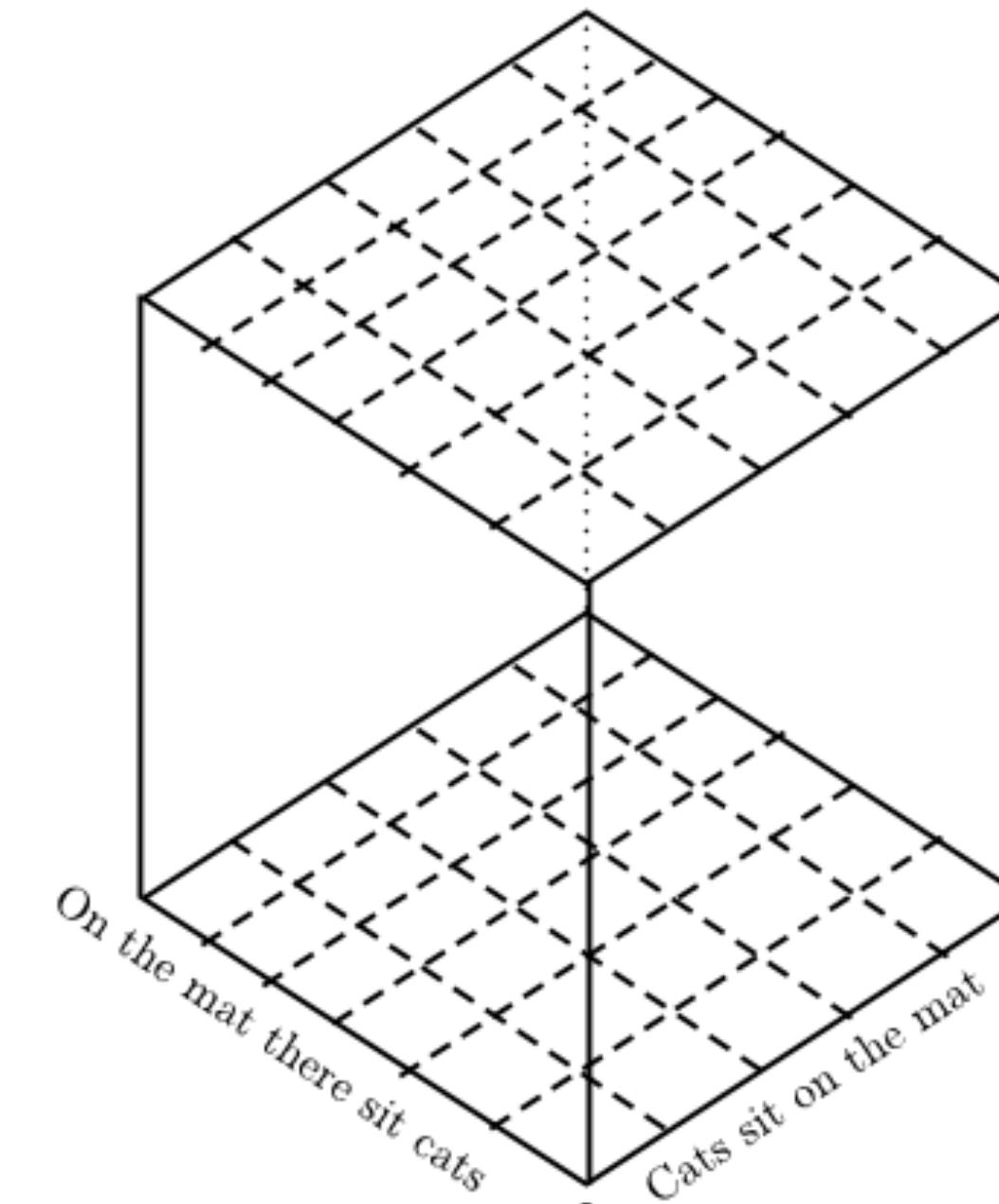


2. Pairwise Word
Interaction Modeling
simCube



DESIGNING VARIOUS [MACHINE LEARNING] MODELS

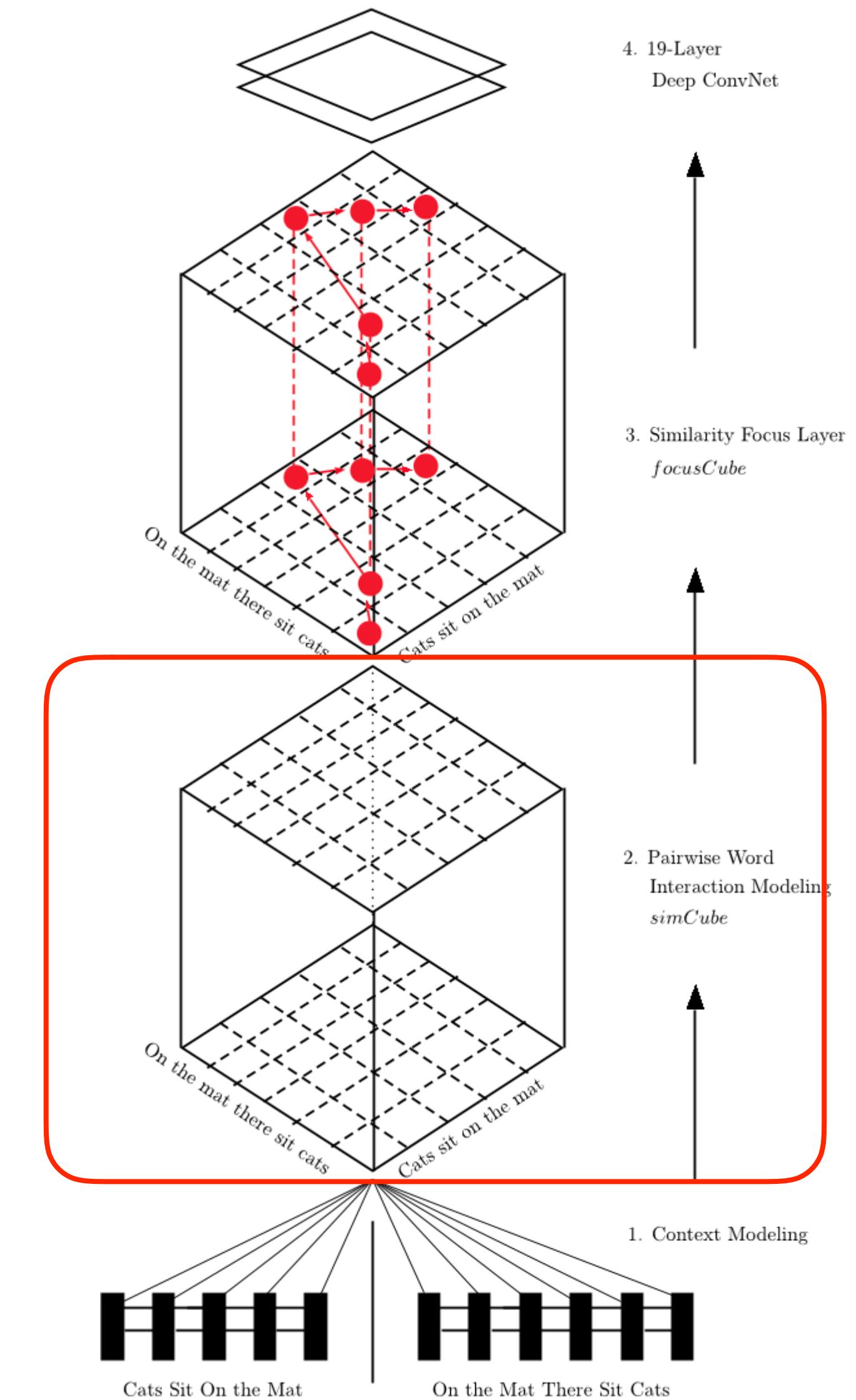
- Deep Pairwise Neural Networks [He et al. 2015; Lan and Xu 2018]



2. Pairwise Word
Interaction Modeling
simCube

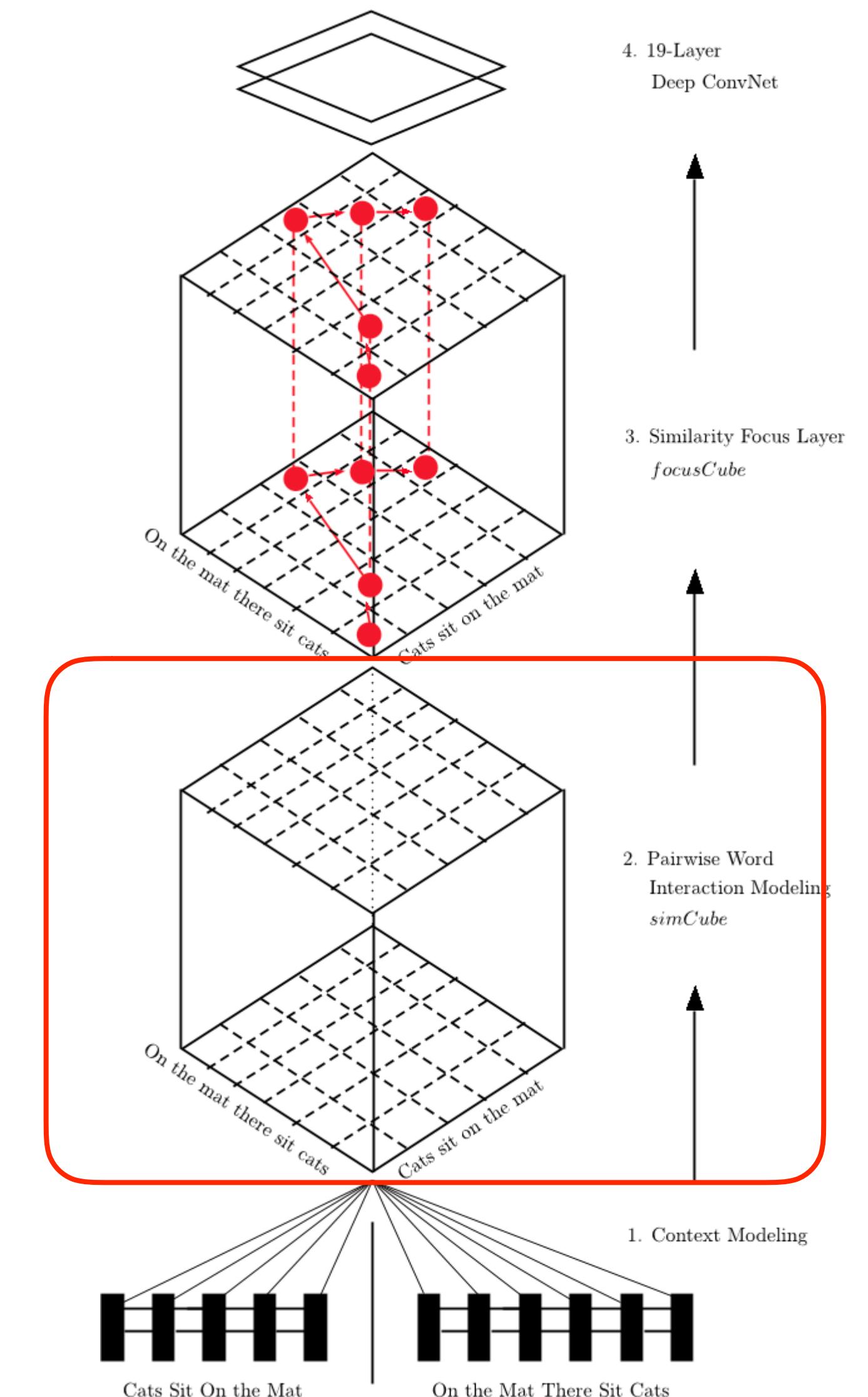
$$coU(\vec{h}_1, \vec{h}_2) = \{\cos(\vec{h}_1, \vec{h}_2), L_2 Euclid(\vec{h}_1, \vec{h}_2), DotProduct(\vec{h}_1, \vec{h}_2)\}$$

Multiple vector similarity measurements used to capture word pair relationship



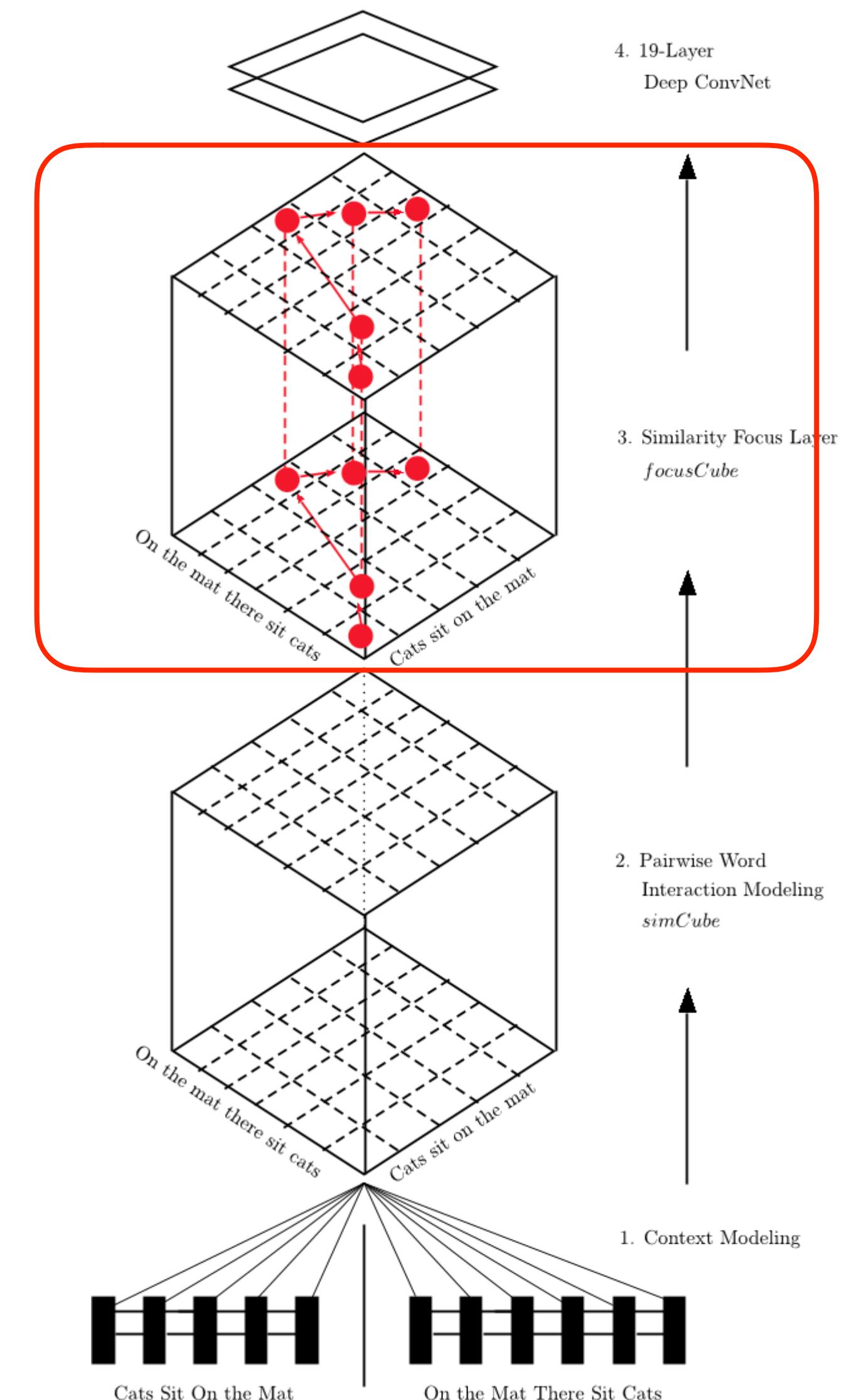
DESIGNING VARIOUS [MACHINE LEARNING] MODELS

- Deep Pairwise Neural Networks [He et al. 2015; Lan and Xu 2018]



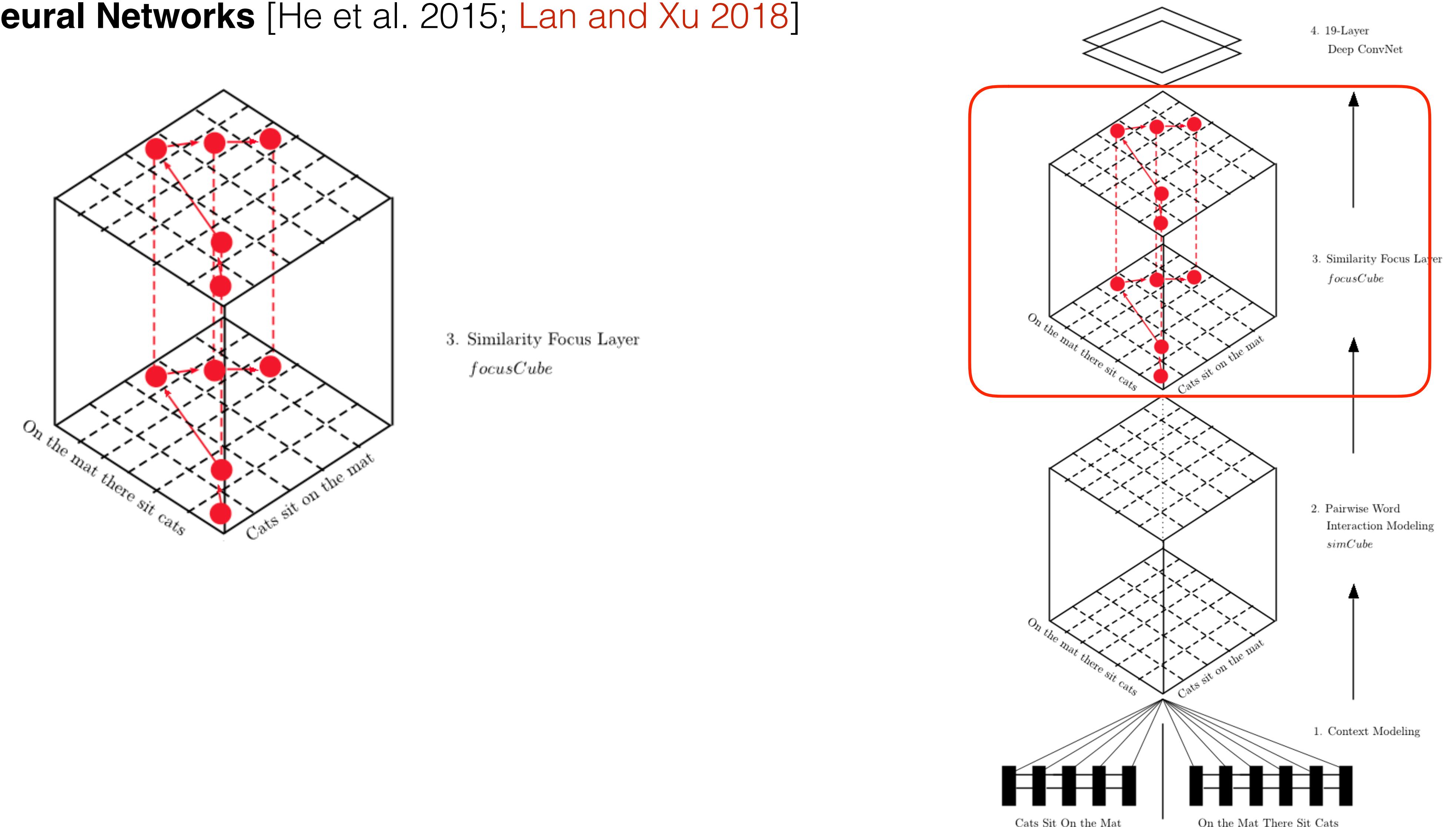
DESIGNING VARIOUS [MACHINE LEARNING] MODELS

- Deep Pairwise Neural Networks [He et al. 2015; Lan and Xu 2018]



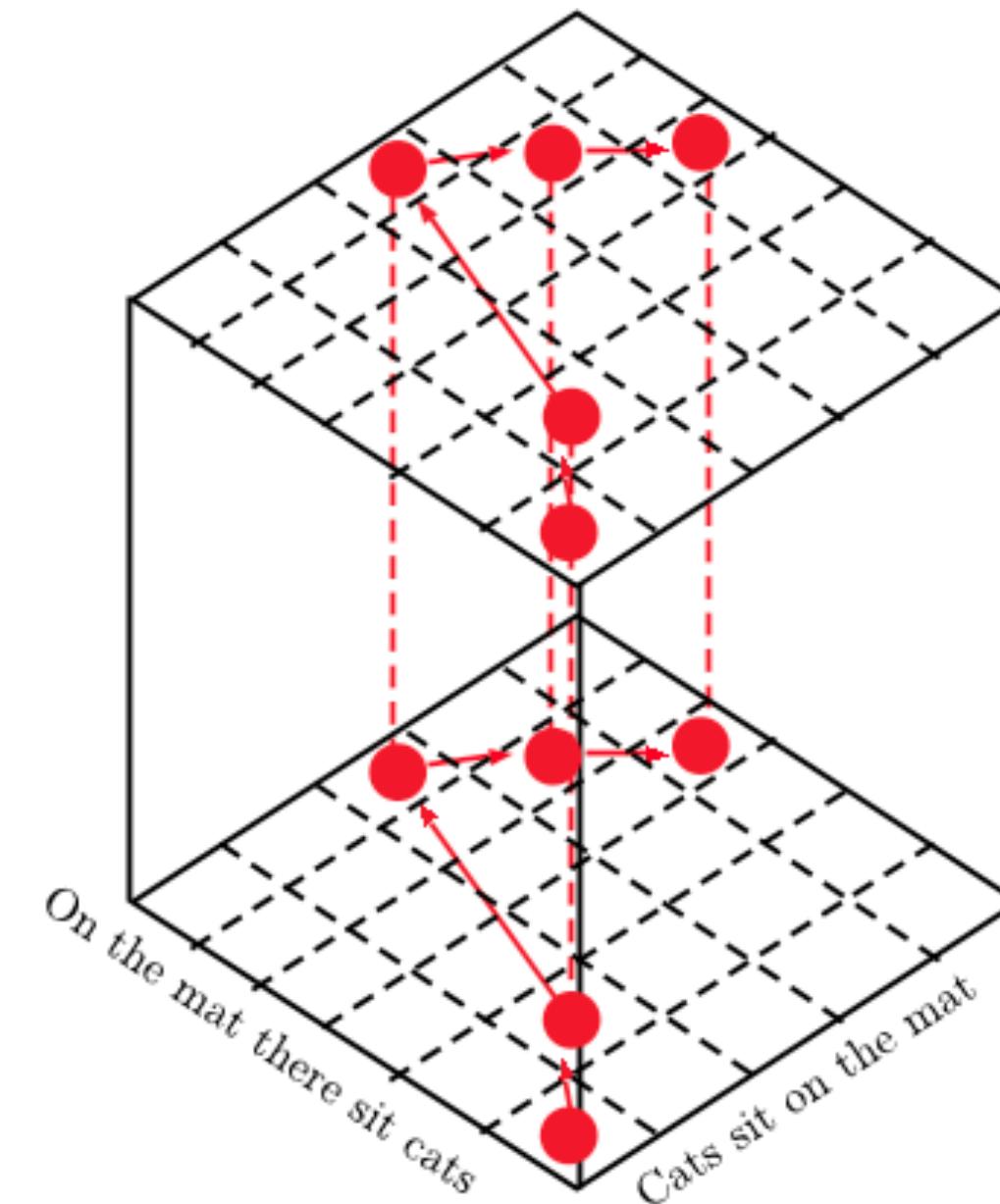
DESIGNING VARIOUS [MACHINE LEARNING] MODELS

- Deep Pairwise Neural Networks [He et al. 2015; Lan and Xu 2018]



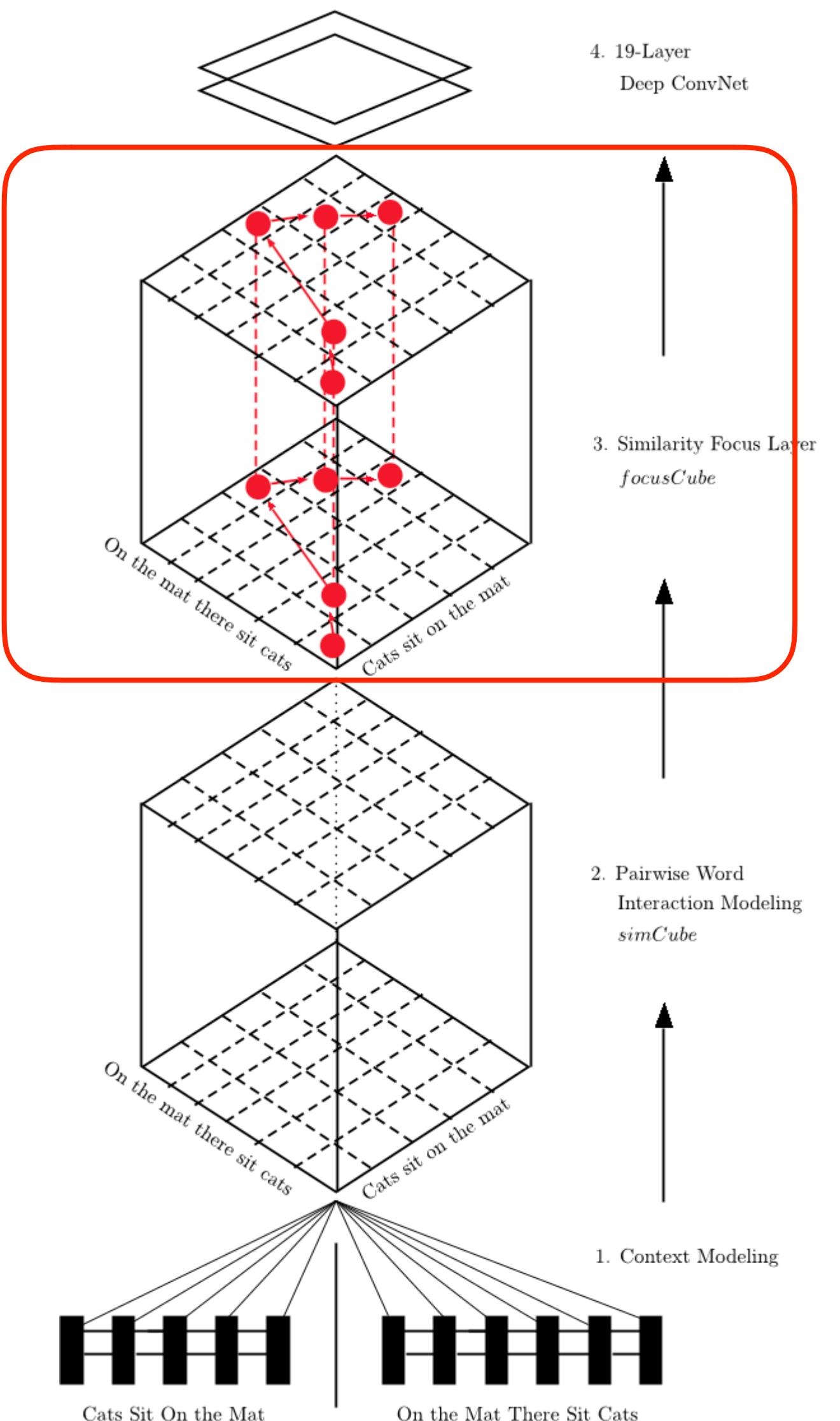
DESIGNING VARIOUS [MACHINE LEARNING] MODELS

- Deep Pairwise Neural Networks [He et al. 2015; Lan and Xu 2018]



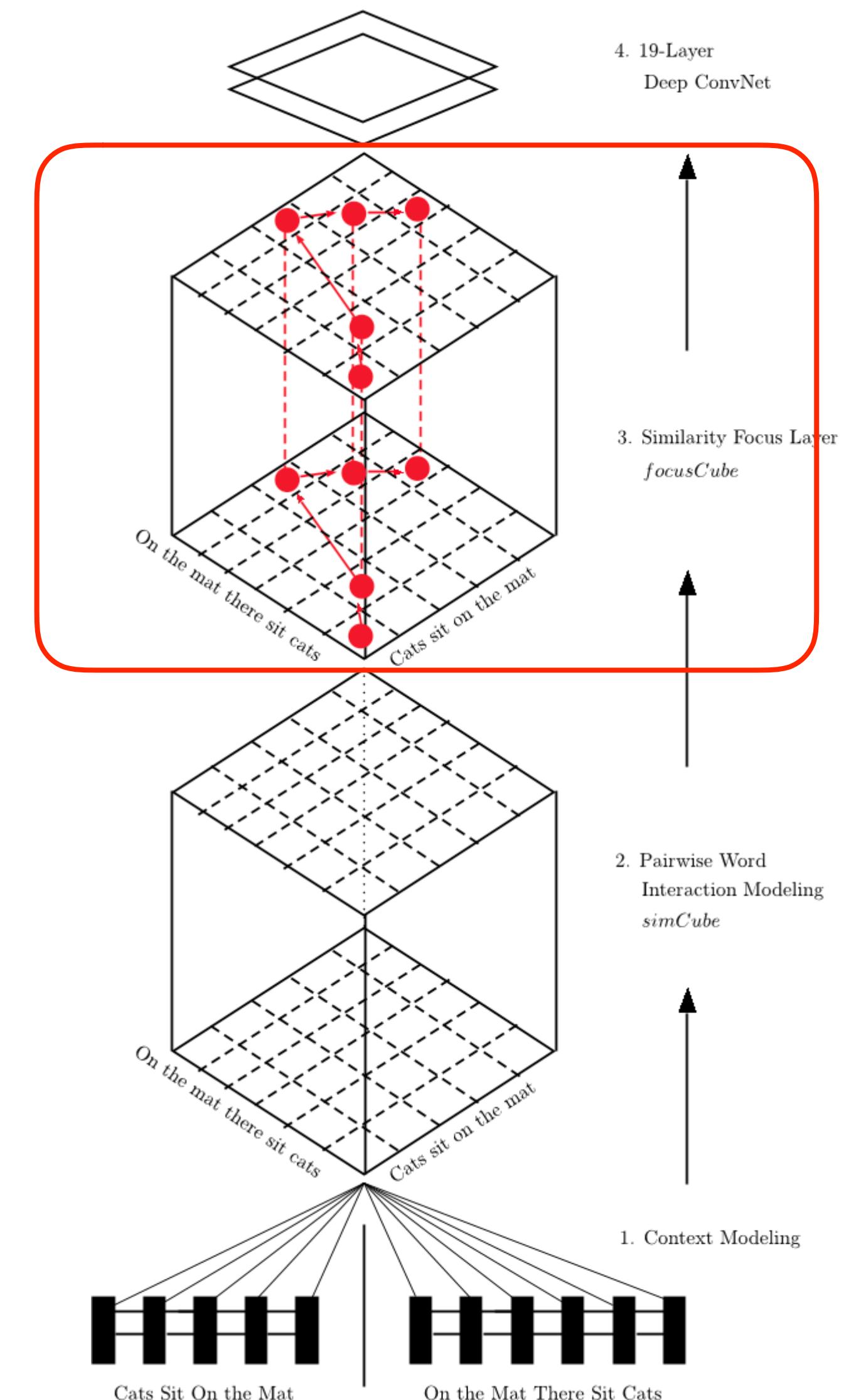
3. Similarity Focus Layer
focusCube

More attention added to top ranked word pairs.



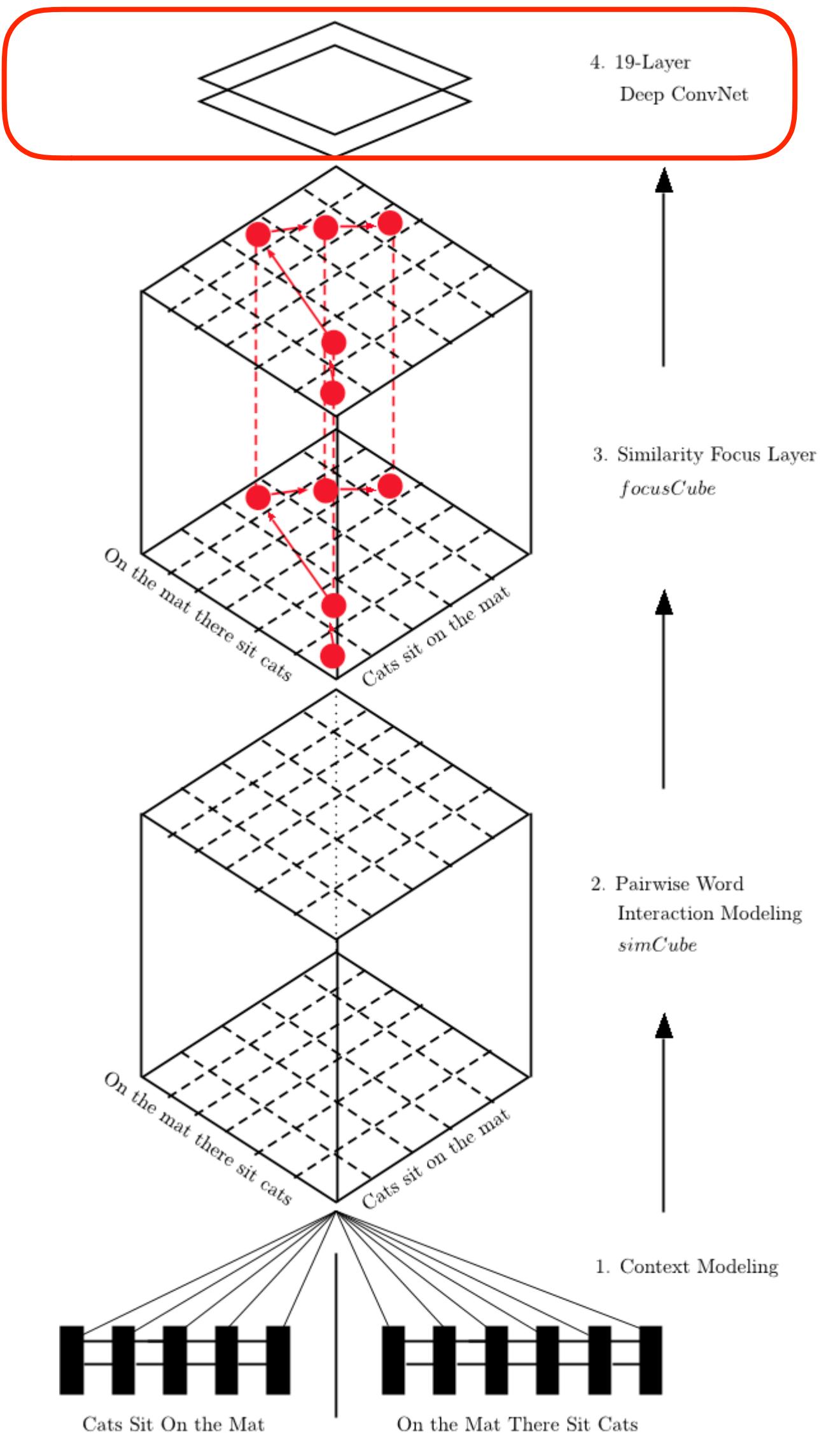
DESIGNING VARIOUS [MACHINE LEARNING] MODELS

- Deep Pairwise Neural Networks [He et al. 2015; Lan and Xu 2018]



DESIGNING VARIOUS [MACHINE LEARNING] MODELS

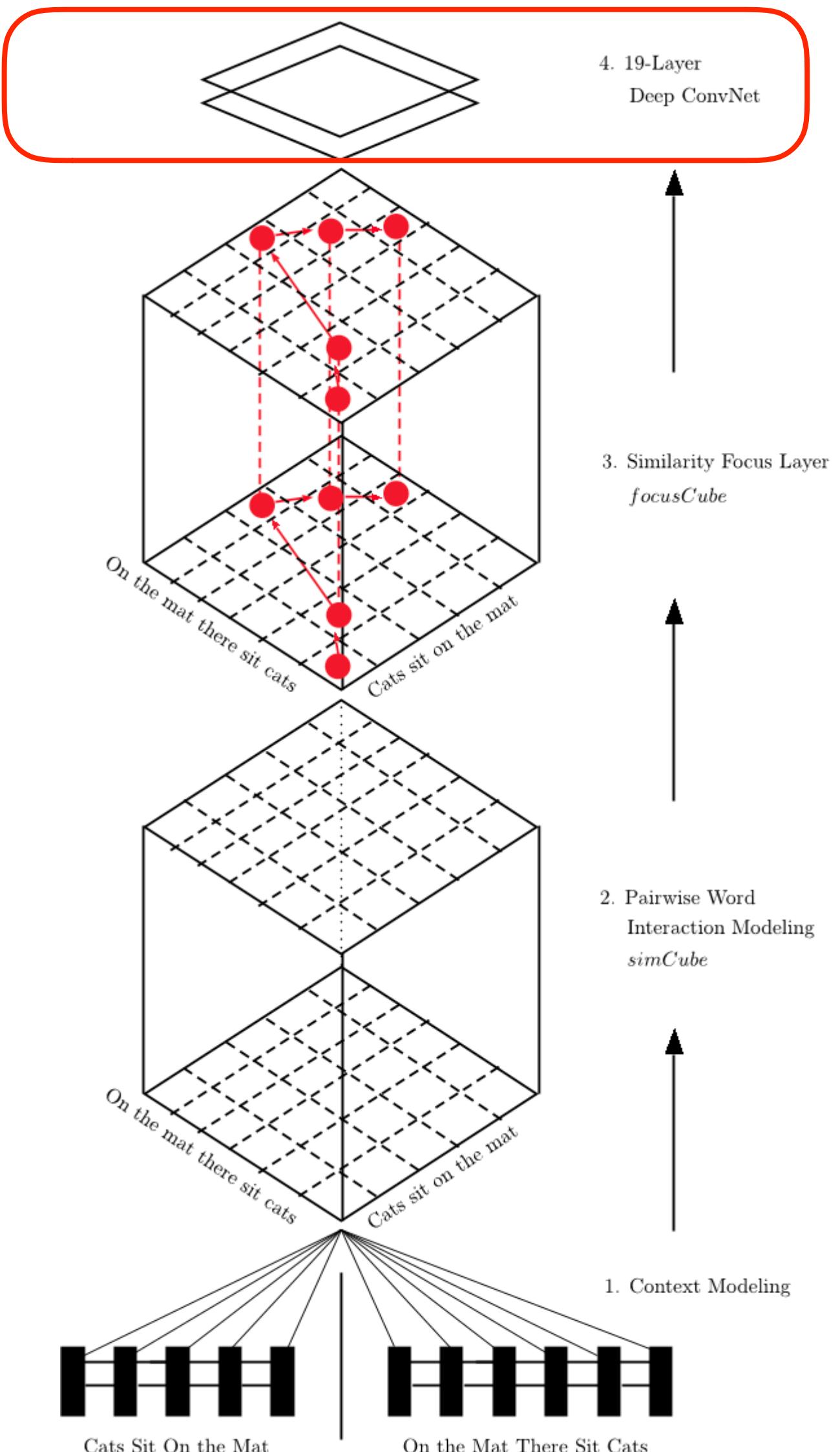
- Deep Pairwise Neural Networks [He et al. 2015; Lan and Xu 2018]



DESIGNING VARIOUS [MACHINE LEARNING] MODELS

- Deep Pairwise Neural Networks [He et al. 2015; Lan and Xu 2018]

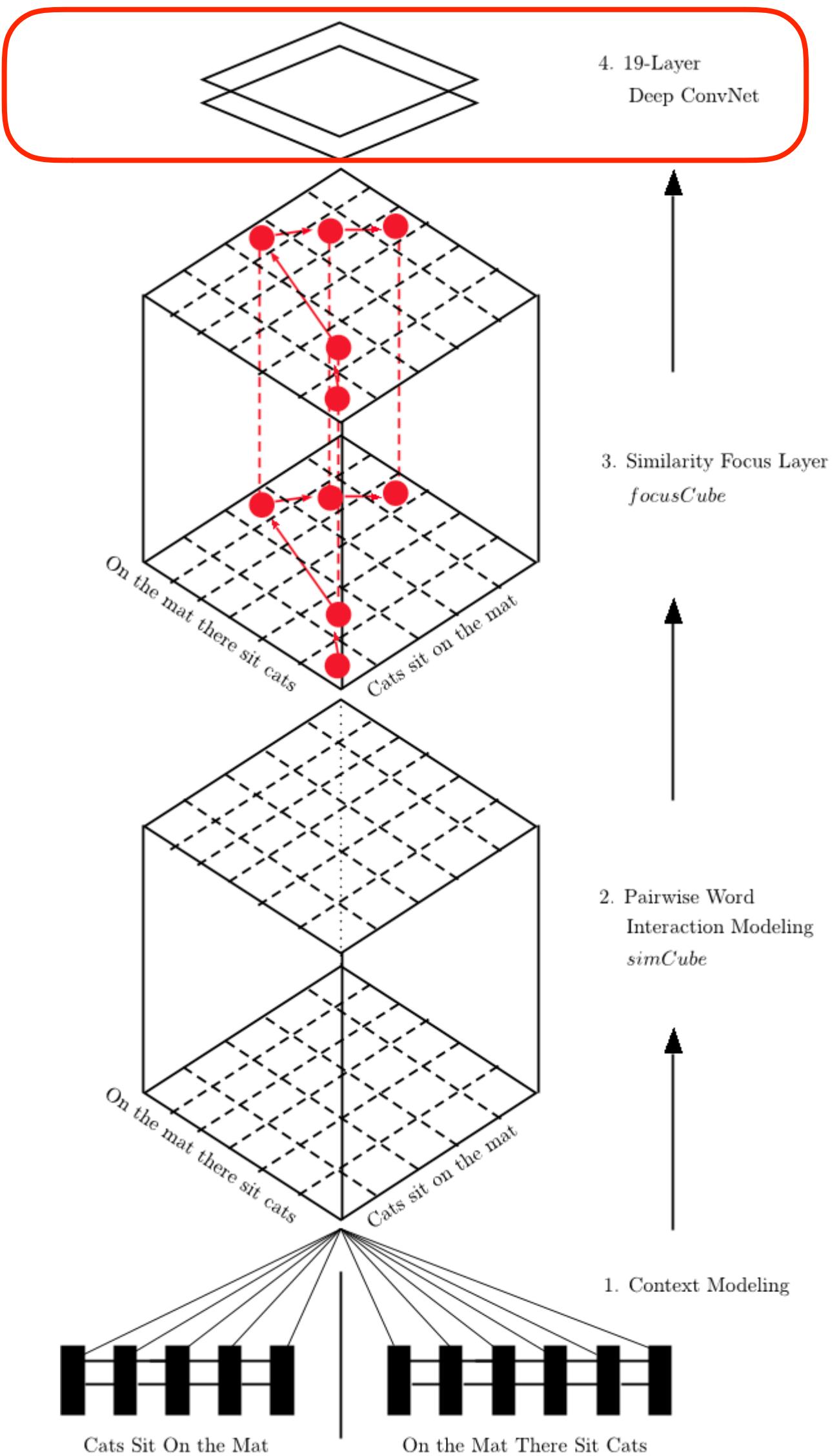
Deep ConvNet Configurations	
Input Size: 32 by 32	Input Size: 48 by 48
Spatial Conv 128: size 3×3 , stride 1, pad 1	
ReLU	
Max Pooling: size 2×2 , stride 2	
Spatial Conv 164: size 3×3 , stride 1, pad 1	
ReLU	
Max Pooling: size 2×2 , stride 2	
Spatial Conv 192: size 3×3 , stride 1, pad 1	
ReLU	
Max Pooling: size 2×2 , stride 2	
Spatial Conv 192: size 3×3 , stride 1, pad 1	
ReLU	
Max Pooling: size 2×2 , stride 2	
Spatial Conv 128: size 3×3 , stride 1, pad 1	
ReLU	
Max Pooling: 2×2 , s2	Max Pooling: 3×3 , s1
Fully-Connected Layer	
ReLU	
Fully-Connected Layer	
LogSoftMax	



DESIGNING VARIOUS [MACHINE LEARNING] MODELS

- Deep Pairwise Neural Networks [He et al. 2015; Lan and Xu 2018]

Deep ConvNet Configurations	
Input Size: 32 by 32	Input Size: 48 by 48
Spatial Conv 128: size 3×3 , stride 1, pad 1	
ReLU	
Max Pooling: size 2×2 , stride 2	
Spatial Conv 164: size 3×3 , stride 1, pad 1	
ReLU	
Max Pooling: size 2×2 , stride 2	
Spatial Conv 192: size 3×3 , stride 1, pad 1	
ReLU	
Max Pooling: size 2×2 , stride 2	
Spatial Conv 192: size 3×3 , stride 1, pad 1	
ReLU	
Max Pooling: size 2×2 , stride 2	
Spatial Conv 128: size 3×3 , stride 1, pad 1	
ReLU	
Max Pooling: 2×2 , s2	Max Pooling: 3×3 , s1
Fully-Connected Layer	
ReLU	
Fully-Connected Layer	
LogSoftMax	

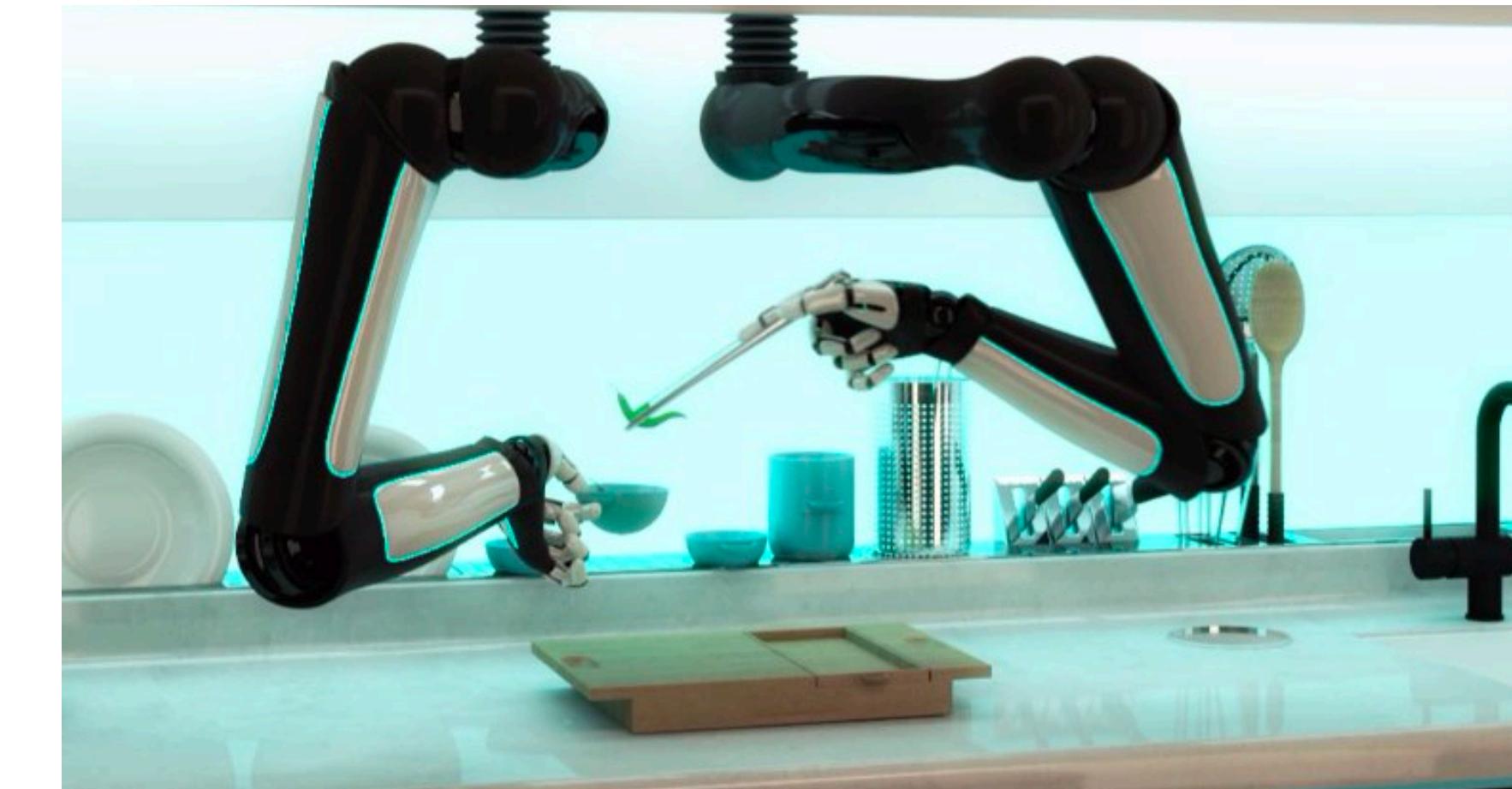


Sentence pair relationship identified by pattern recognition through ConvNet.

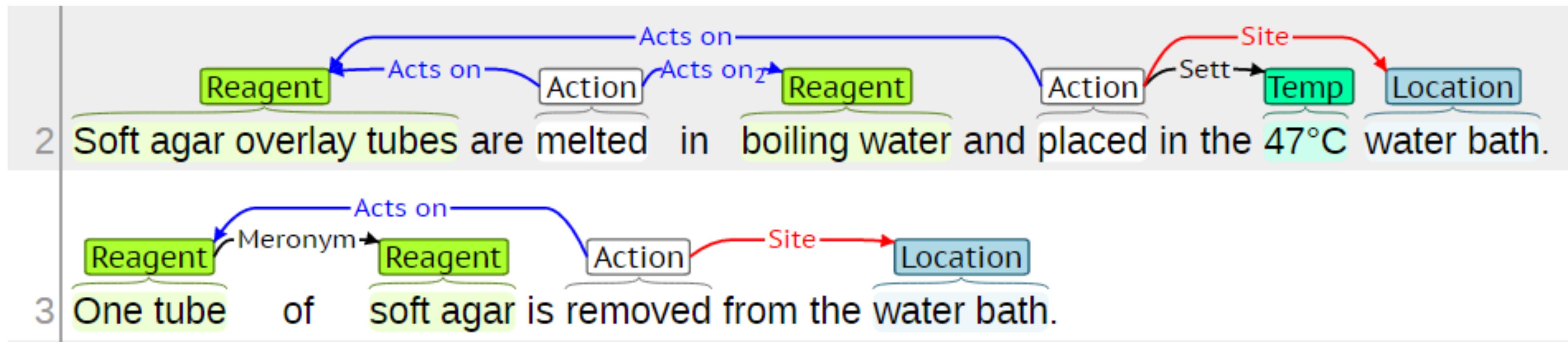
[DATA SCIENCE] FOR LINGUISTIC STYLES



MACHINE READING OF INSTRUCTIONS



Wet Lab Protocols:



Assistant Professor Wei Xu

Department of Computer Science and Engineering
Ohio State University

<http://web.cse.ohio-state.edu/~weixu/>

THANK YOU

