Numeracy for Language Models:

Evaluating and Improving their Ability to Predict Numbers

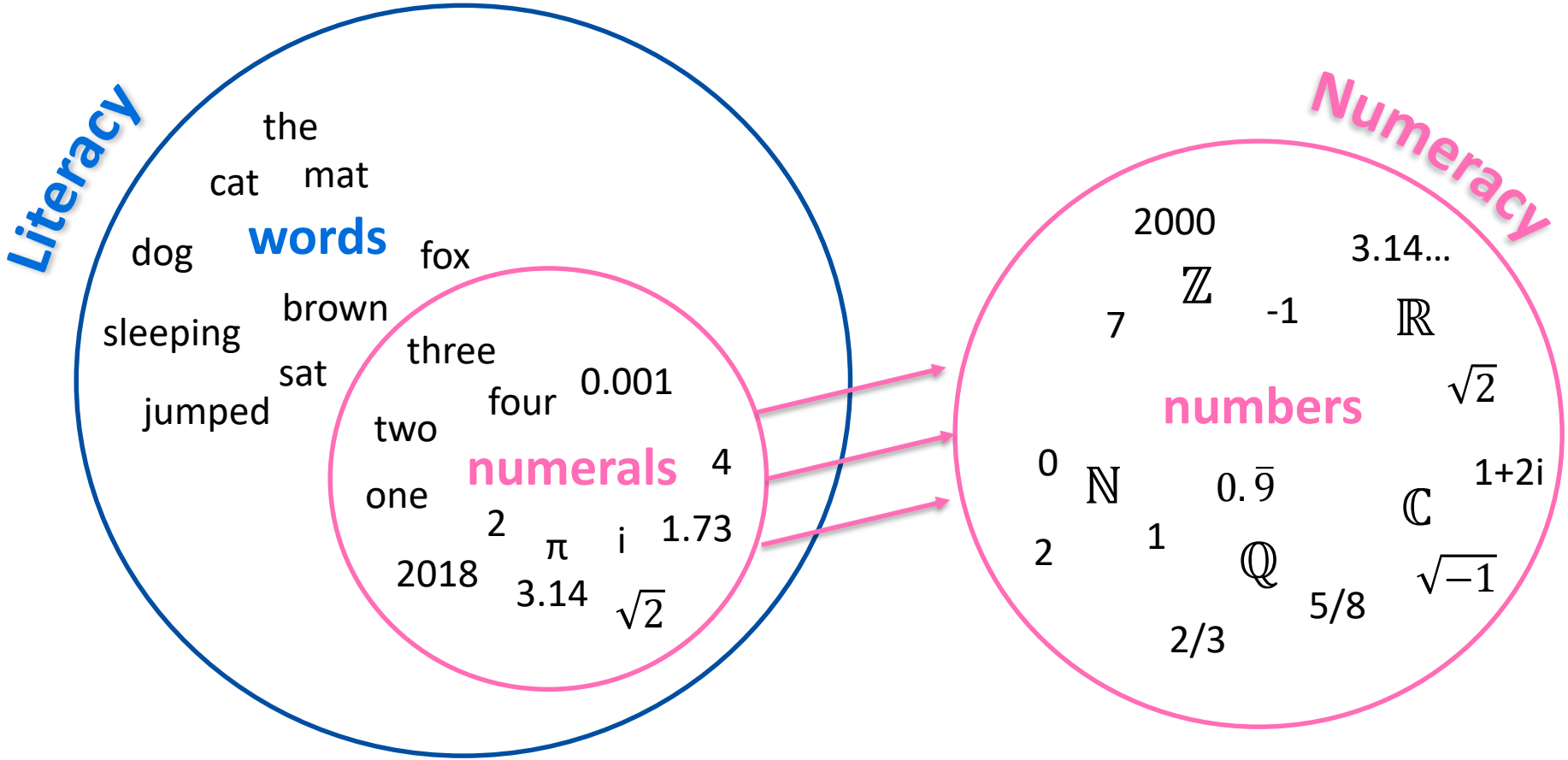
Georgios Spithourakis, Steffen Petersen, Sebastian Riedel

Machine Reading Group

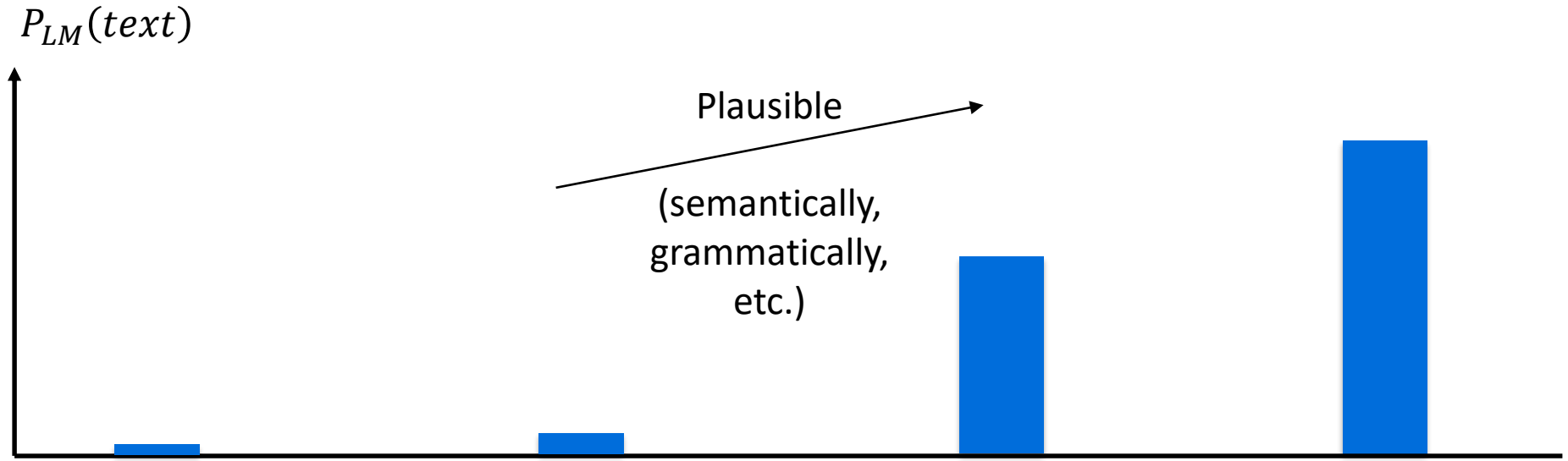
UCL

Farr The Farr Institute of Health Informatics Research

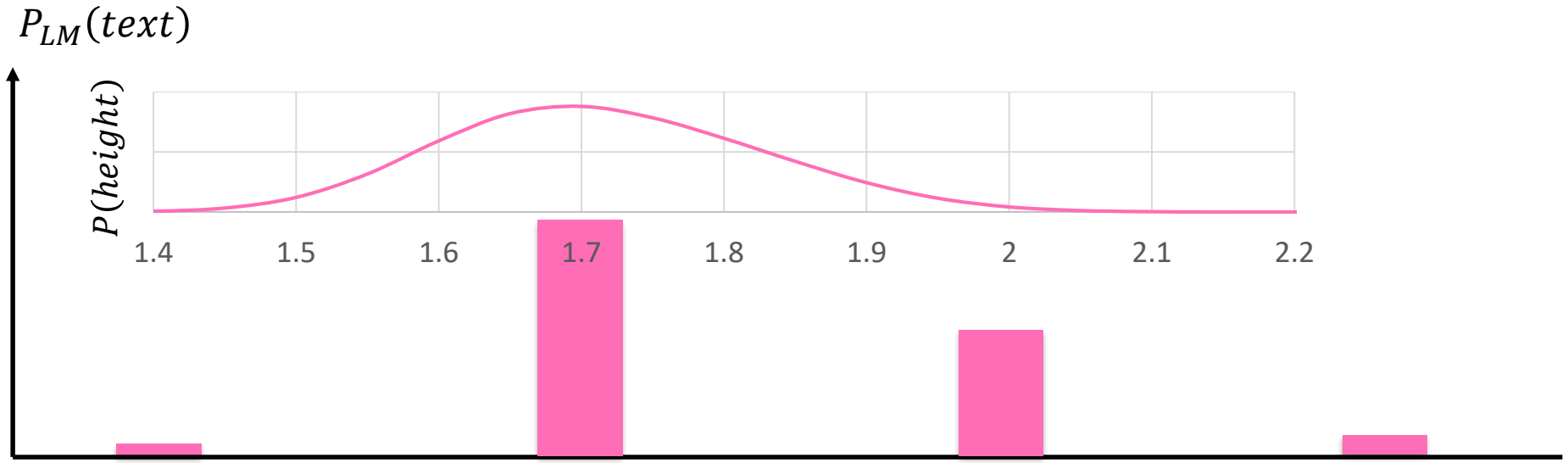
Numeracy


Literacy
the
cat
mat
words
dog
fox
sleeping
brown
sat
jumped
three
four
0.001
two
numerals
4
one
2
π
i
1.73
2018
3.14
$\sqrt{2}$
Numeracy
2000
3.14…
$\mathbb{Z}$
7
-1
$\mathbb{R}$
$\sqrt{2}$
numbers
0
$\mathbb{N}$
$0.\bar{9}$
1+2i
$\mathbb{C}$
2
1
$\mathbb{Q}$
$\sqrt{-1}$
5/8
2/3

Literate Language Models



Numerate Language Models



‘John is **0** m tall’ ‘John is **1.7** m tall’ ‘John is **2** m tall’ ‘John is **999** m tall’

Numeracy Matters

0.5
‘Unemployment of the US is 5 %’
50
500

0.0
23.2
‘Patient’s temperature is 36.6 degrees’
41.9
98.6

0
‘Our model is 10 times better than the baseline’
100
1000

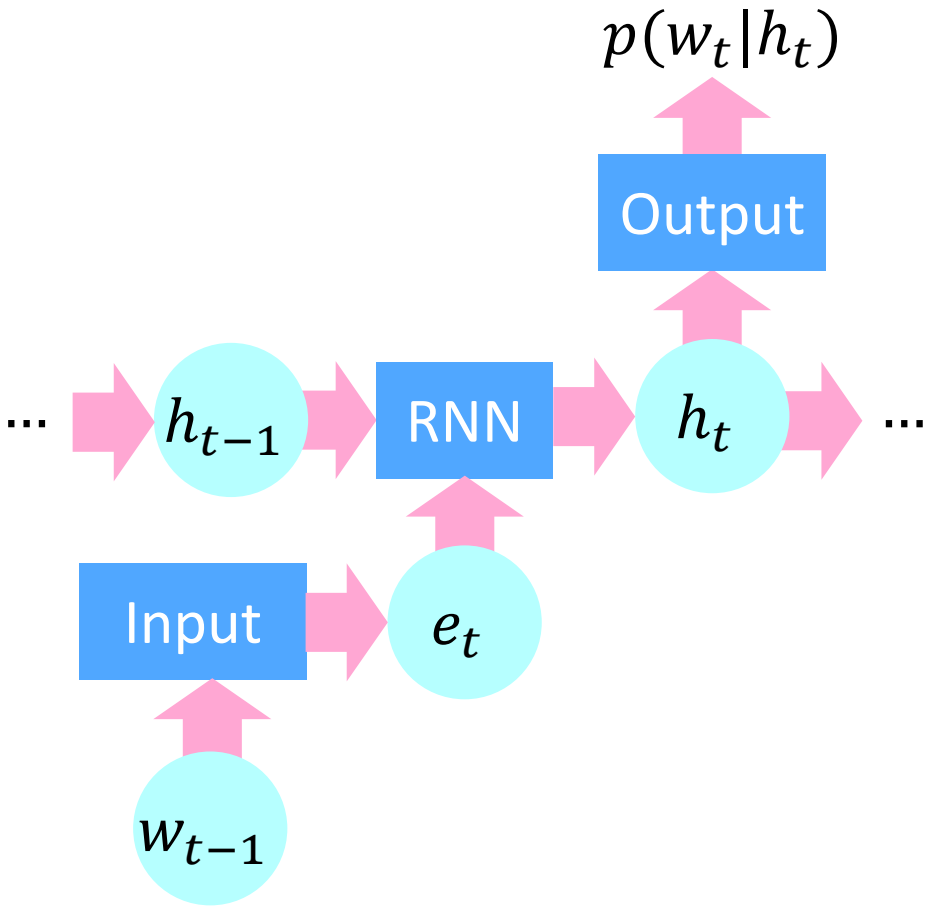
Q1: Are existing LMs numerate?

Q2: How to improve the numeracy of LMs?

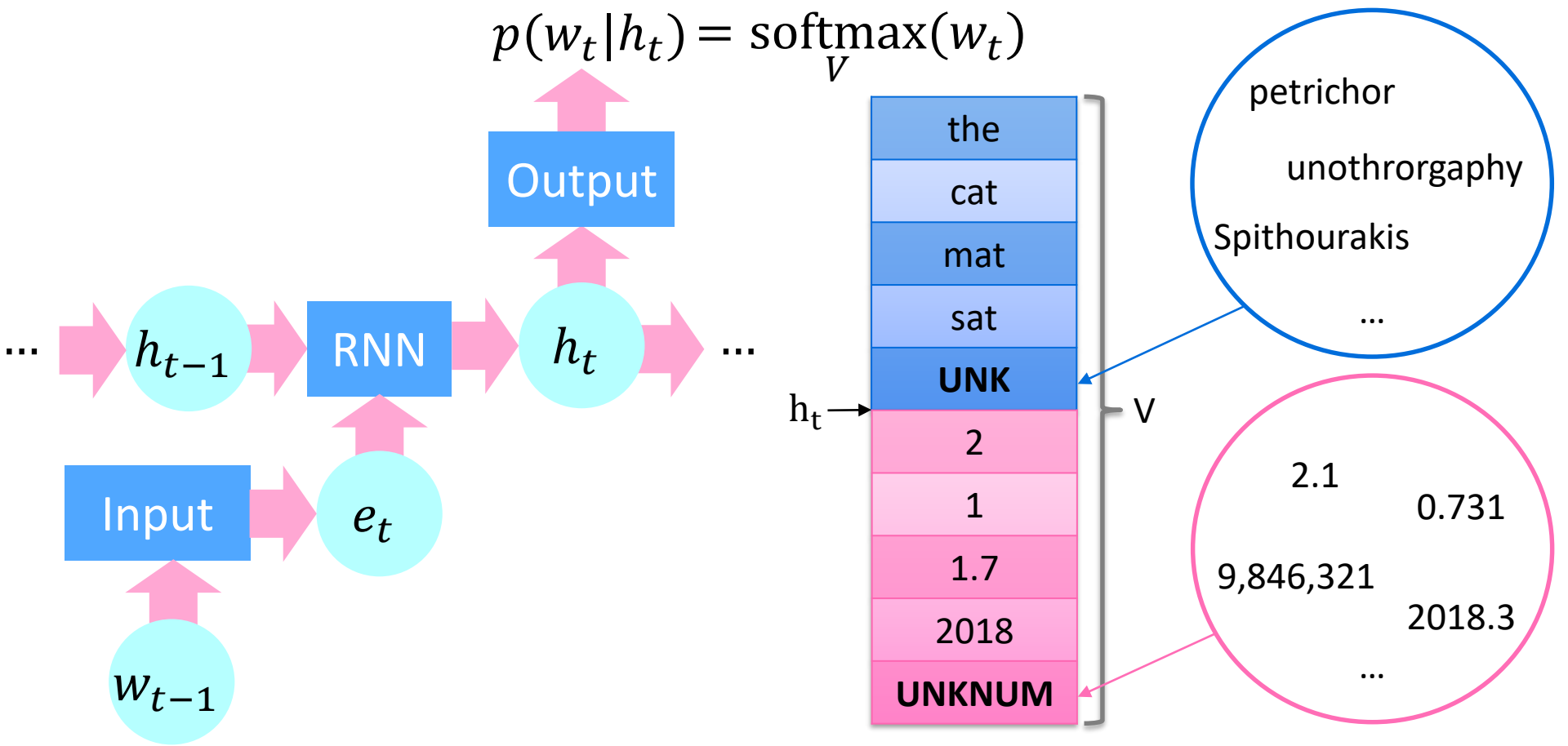
Q1: Are existing LMs numerate?

Q2: How to improve the numeracy of LMs?

A Neural Language Model



A Neural Language Model



Evaluation: Adjusted Perplexity

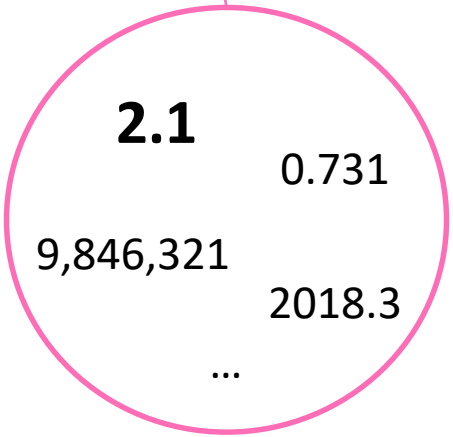
Perplexity

John is **2.1** m tall

$p(2.1) = p($ UNKNUM $)$

BUT

$+$
$p(0.731)$
$+$
$p(9{,}846{,}321)$
$+$
$\vdots$



Evaluation: Adjusted Perplexity

Perplexity

John is **2.1** m tall

$$p(2.1) = p(\ \text{UNKNUM}\ )$$

BUT

$$+\ p(0.731) + p(9{,}846{,}321) + \vdots$$

2.1

0.731

9,846,321

2018.3

…

Adjusted Perplexity [Ueberla, 1994]

$$p(2.1) = \frac{p(\ \text{UNKNUM}\ )}{|w \in \text{UNKNUM}|}$$

from test data

a.k.a. Unknown-Penalised Perplexity

[Ahn et al., 2016]

Datasets

Clinical Dataset

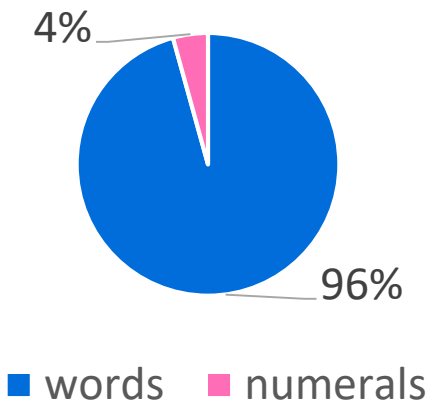
16,015 clinical patient reports

Source: London Chest Hospital

Scientific Dataset

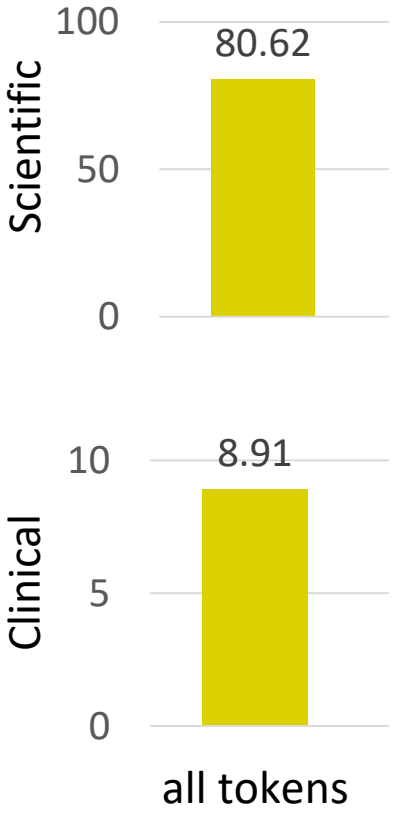
20,962 paragraphs
from scientific papers

Source: ARXIV



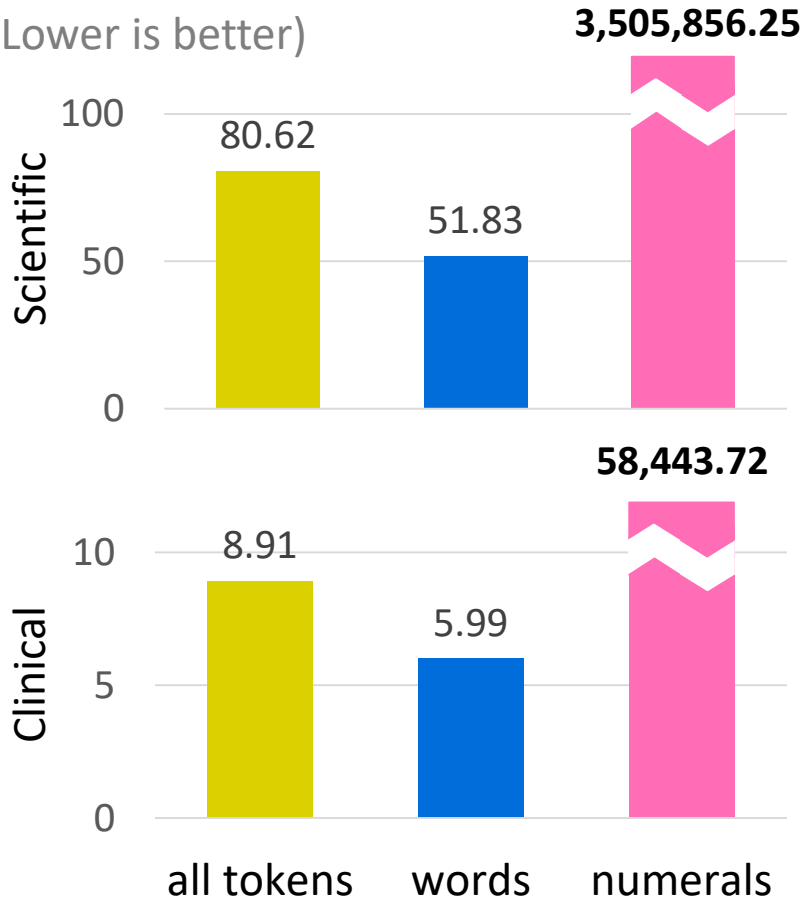
Results: Adjusted Perplexity

(Lower is better)

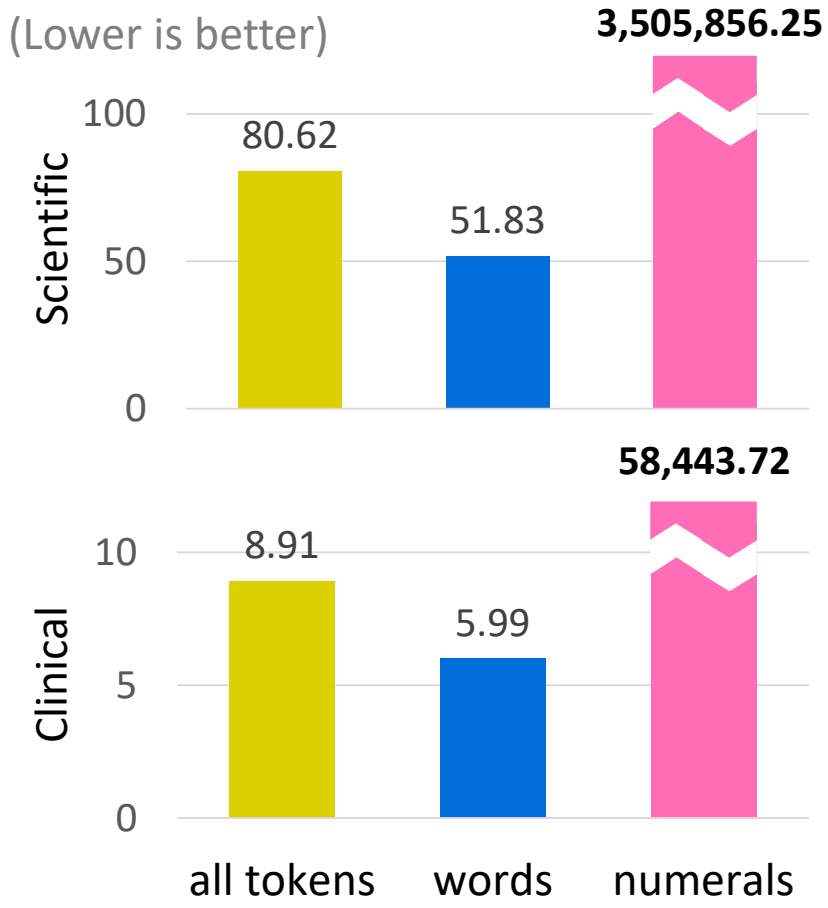


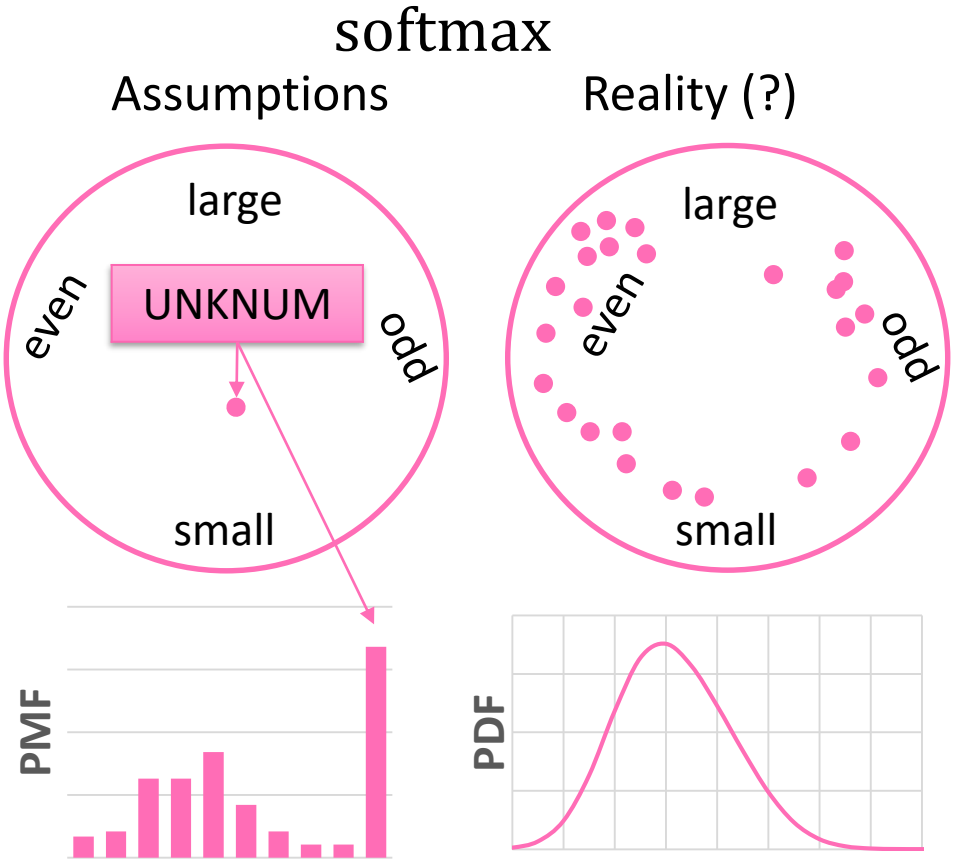
Results: Adjusted Perplexity

(Lower is better)



Results: Adjusted Perplexity

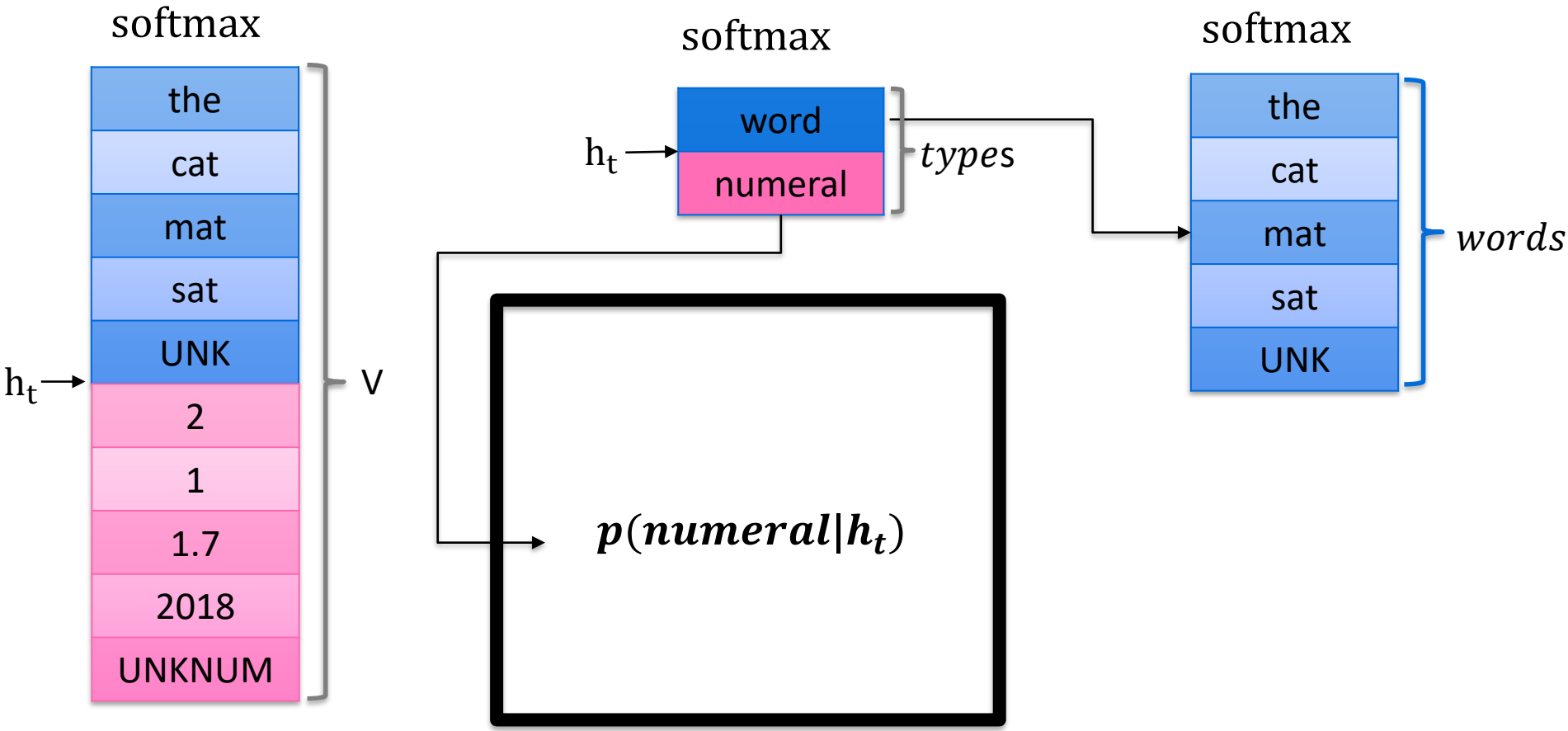




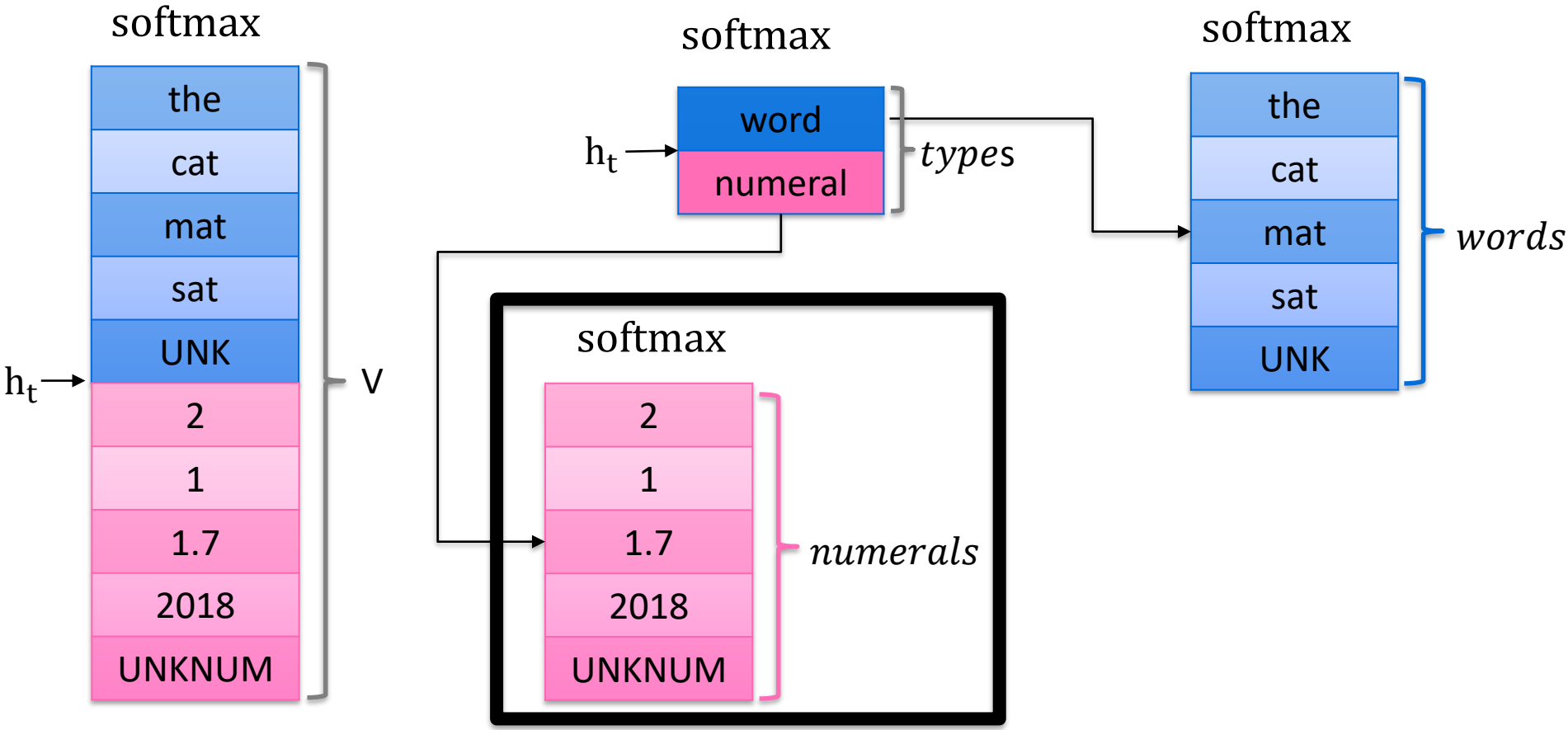
Q1: Are existing LMs numerate?

Q2: How to improve the numeracy of LMs?

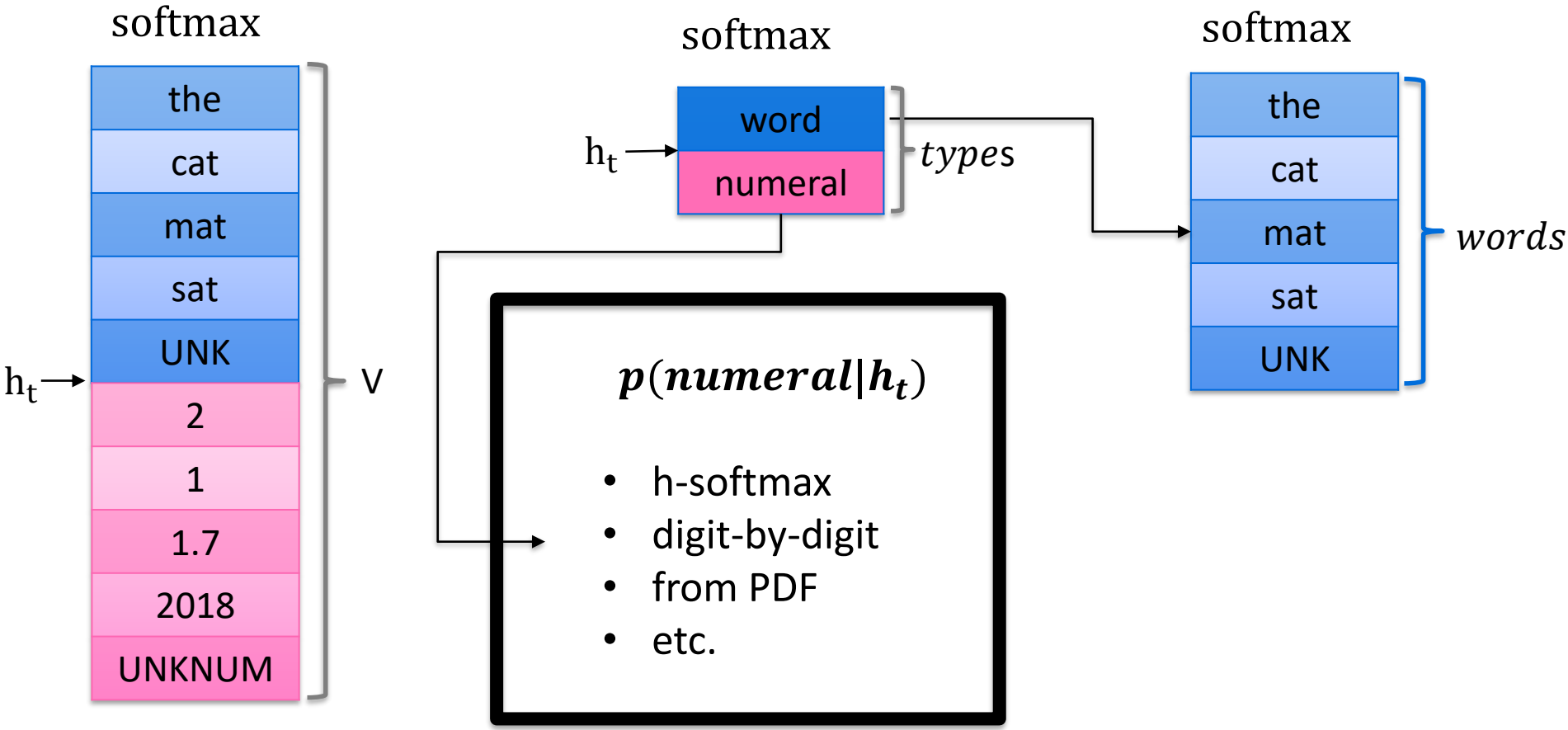
Strategy: Softmax & Hierarchical Softmax



Strategy: Softmax & Hierarchical Softmax

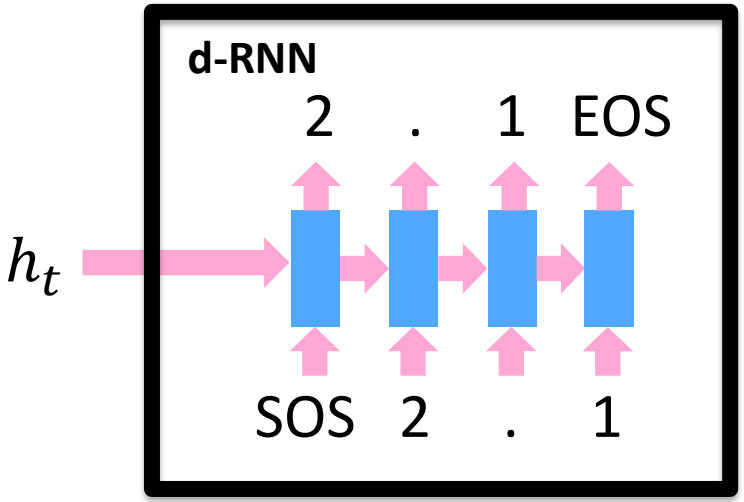


Strategy: Softmax & Hierarchical Softmax



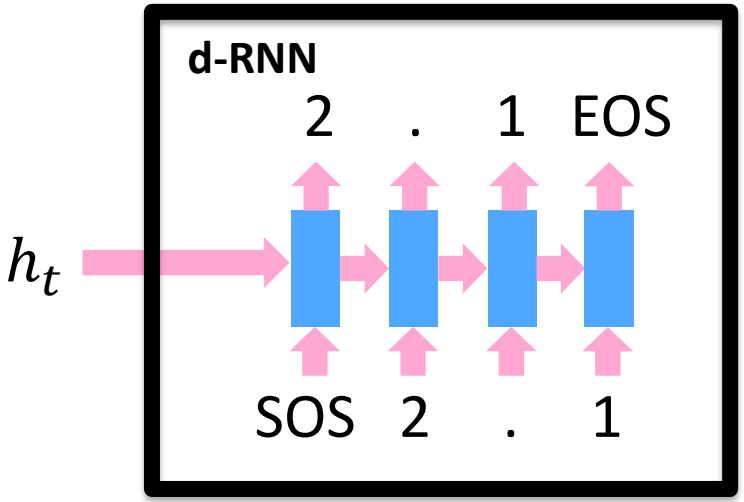
Strategy: Digit-by-Digit Composition

$$p(2.1) = p(2)p(.|2)p(1|2.)p(EOS|2.1)$$



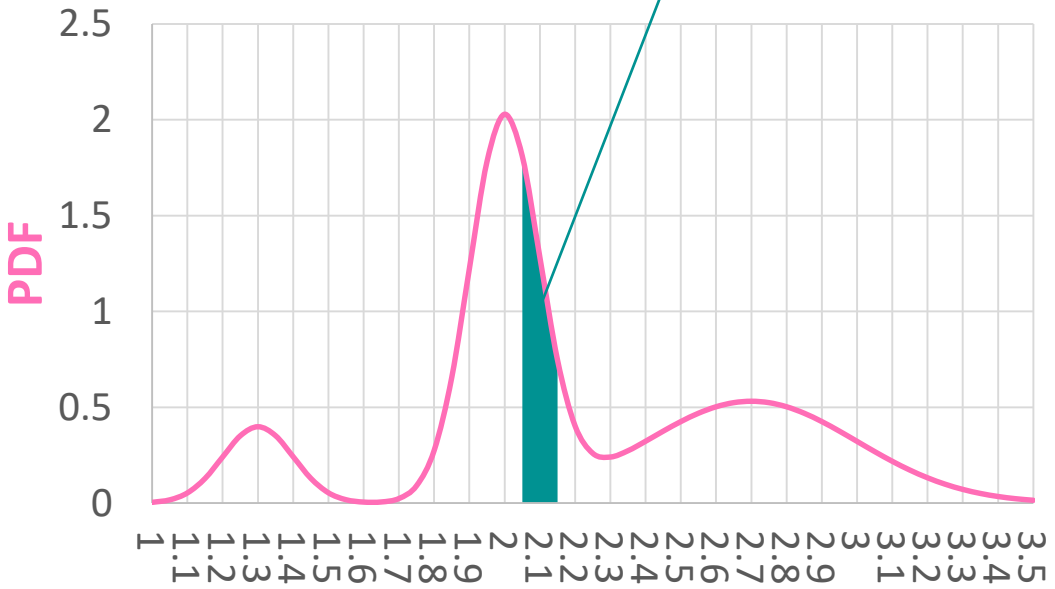
Strategy: Digit-by-Digit Composition

$$p(2.1) = p(2)p(.|2)p(1|2.)p(EOS|2.1)$$



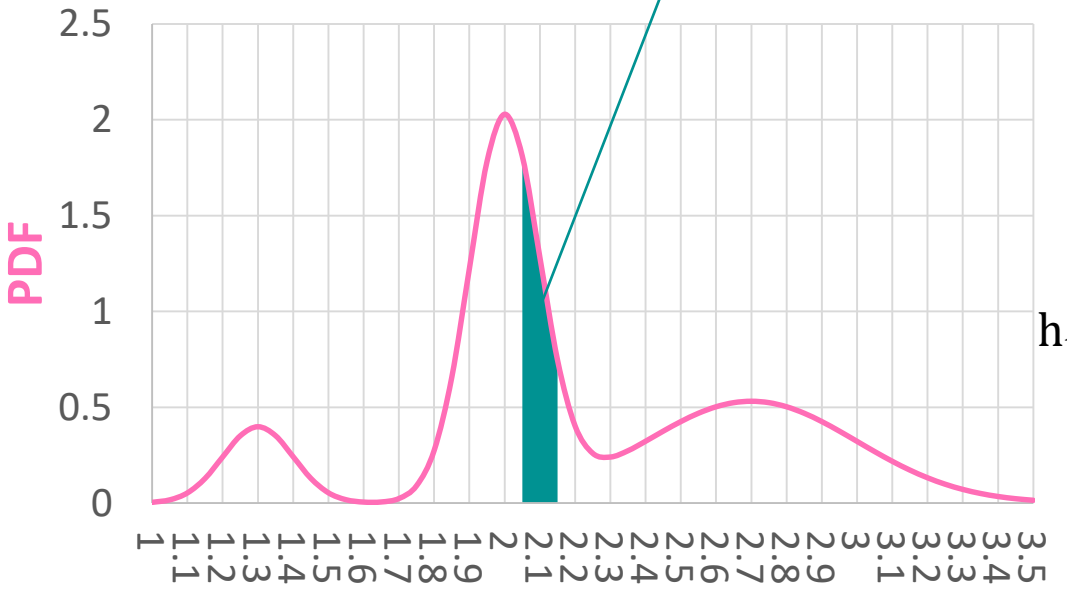
Strategy: from continuous PDF

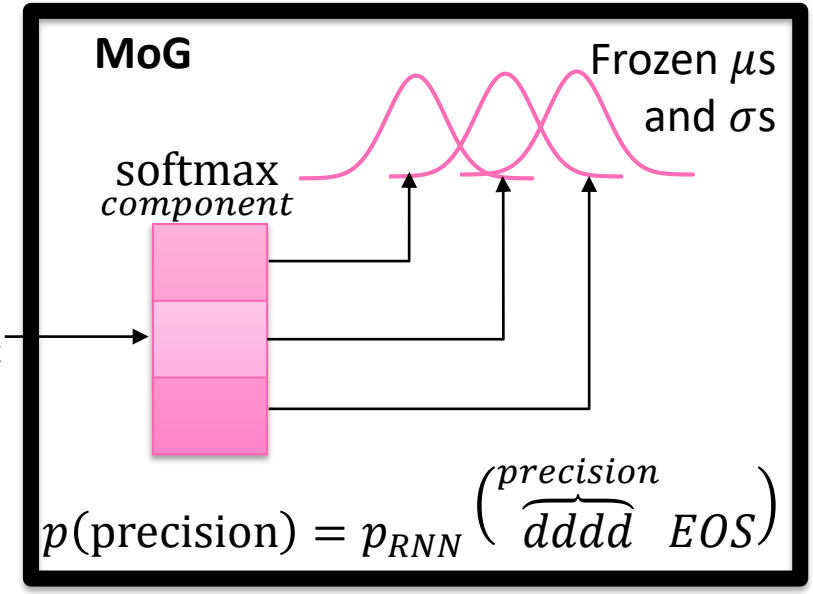
$$p(\textbf{numeral} = 2.1) = p_{PMF}(2.05 < \textbf{number} < 2.15 \,|\text{precision} = 1\,) \times\; p(\text{precision} = 1\,)$$



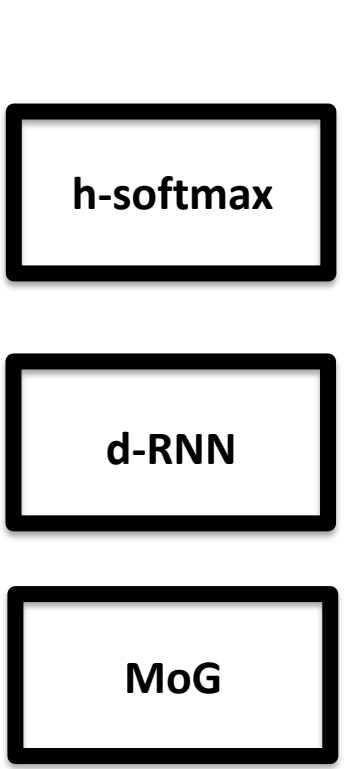
Strategy: from continuous PDF

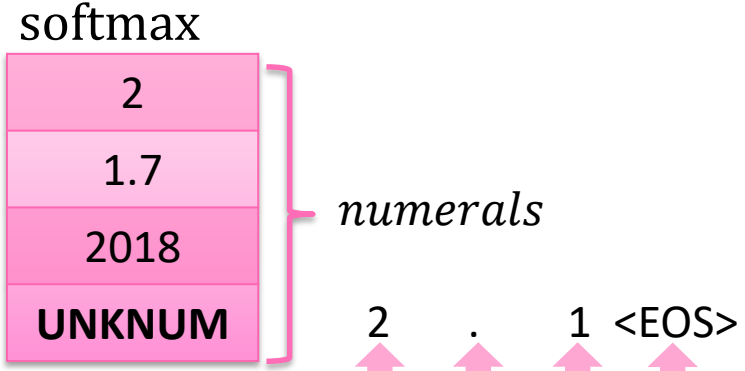
$$p(\textbf{numeral} = 2.1) = p_{PMF}(2.05 < \textbf{number} < 2.15 \,|\text{precision} = 1\,) \times\; p(\text{precision} = 1\,)$$

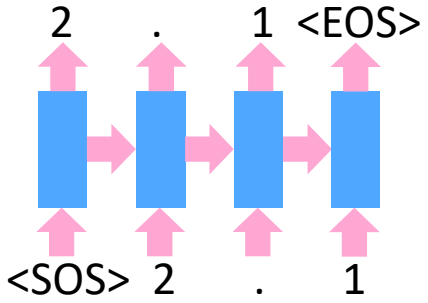


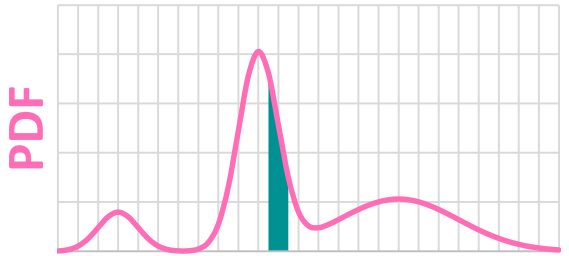


Overview of Strategies

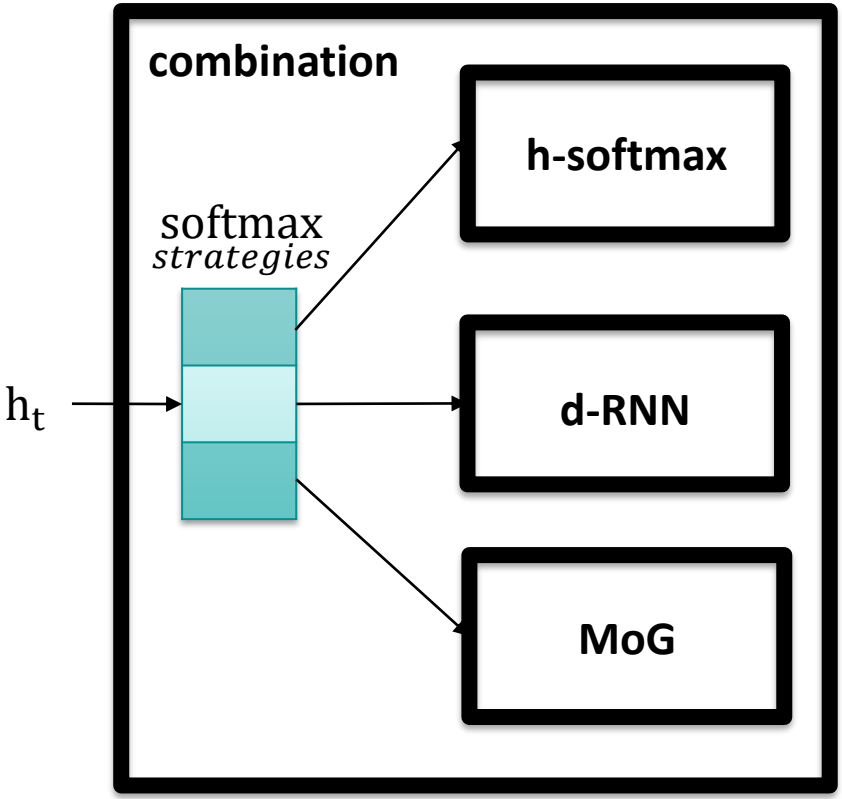


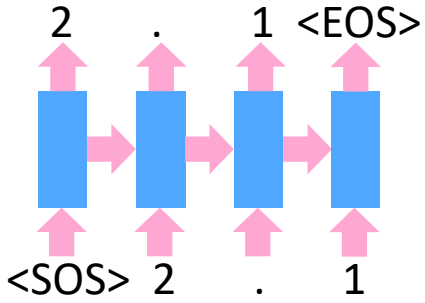


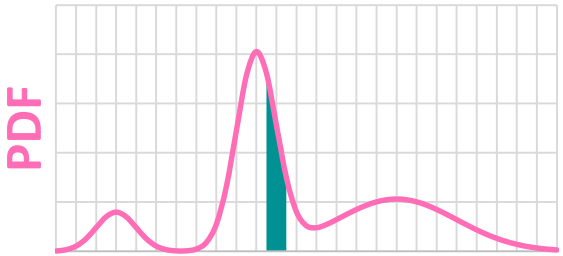




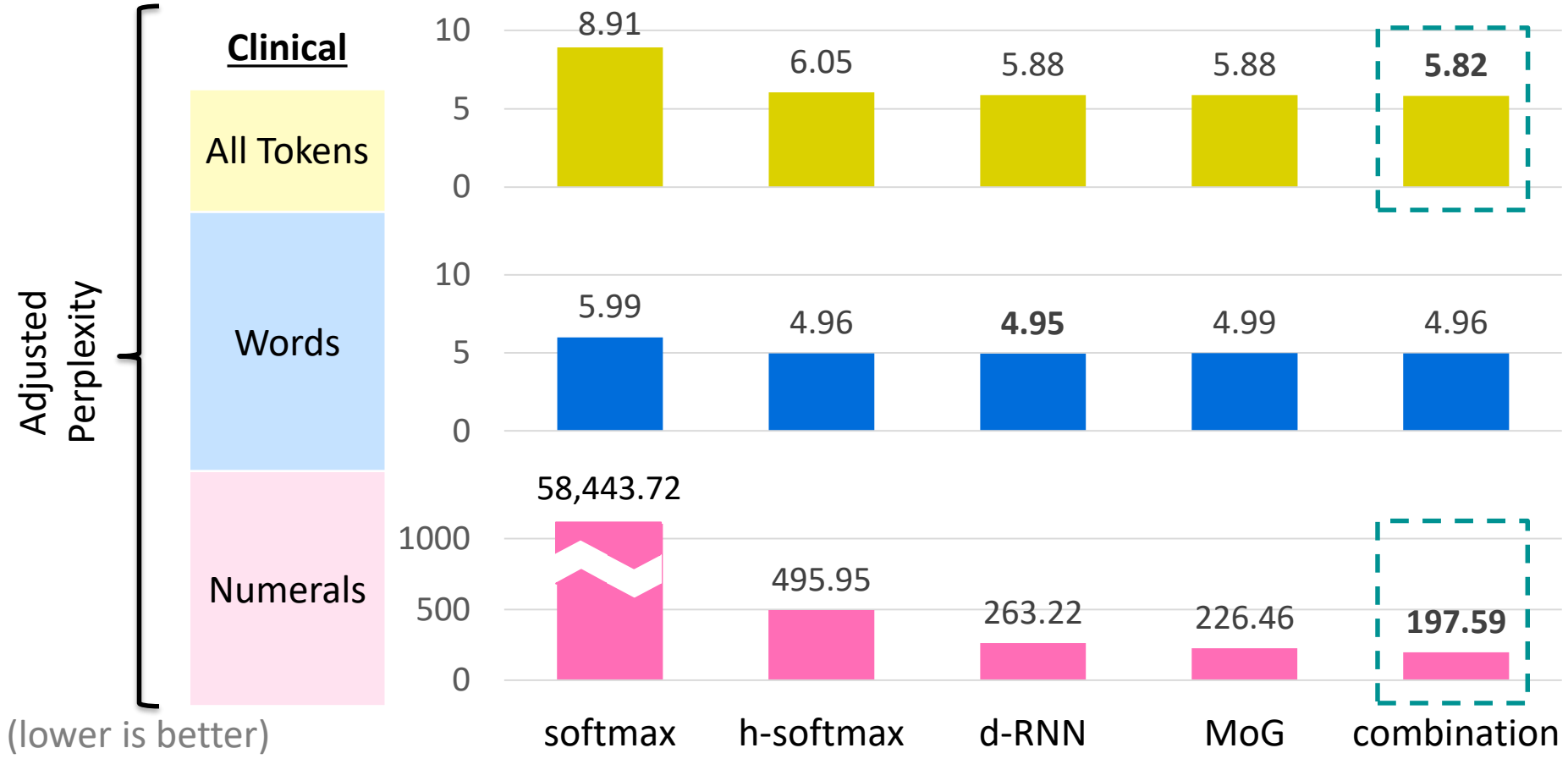
Overview of Strategies



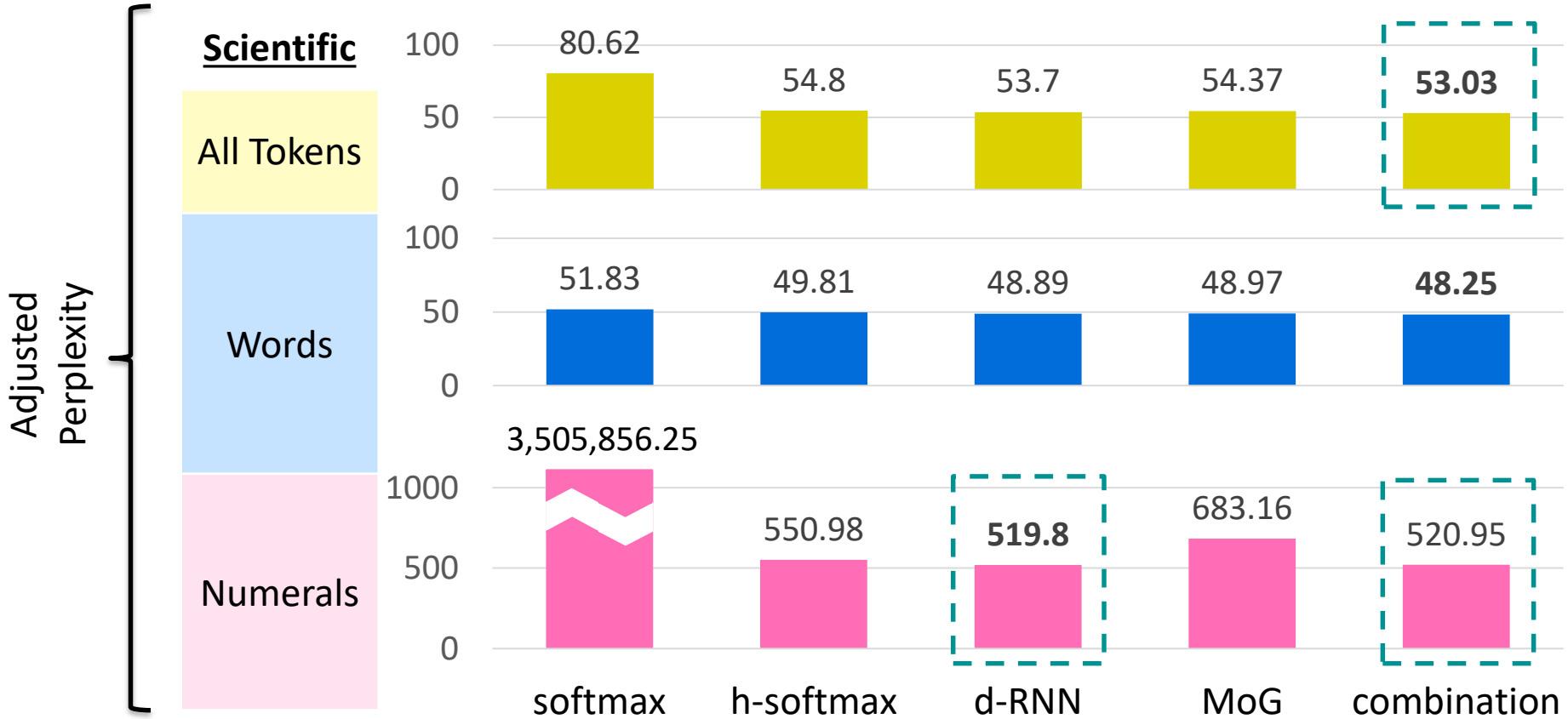




Results: Language Modelling (1)



Results: Language Modelling (2)



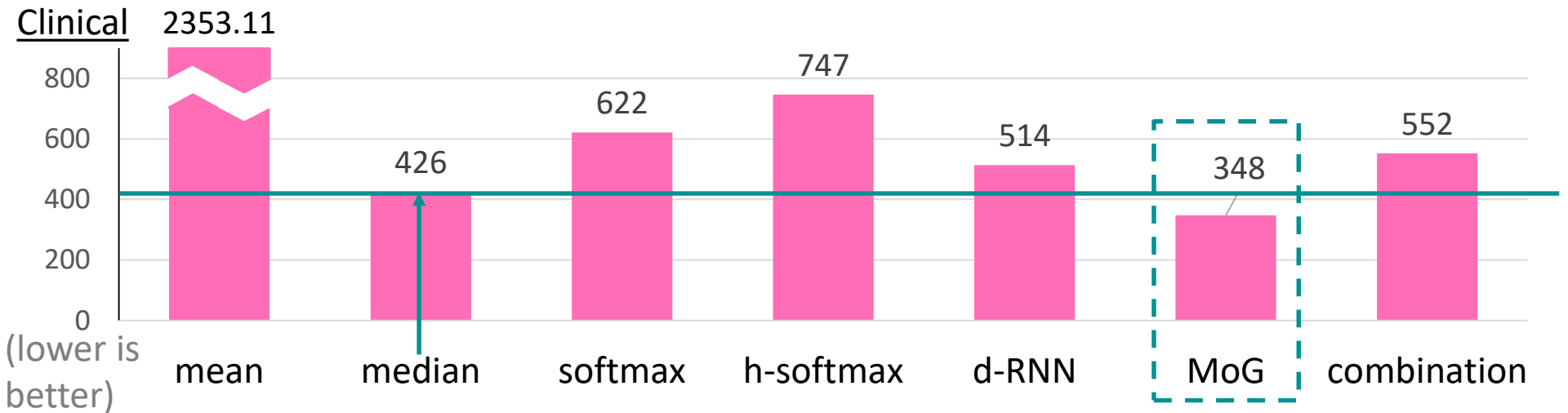
(lower is better)

Results: Number Prediction

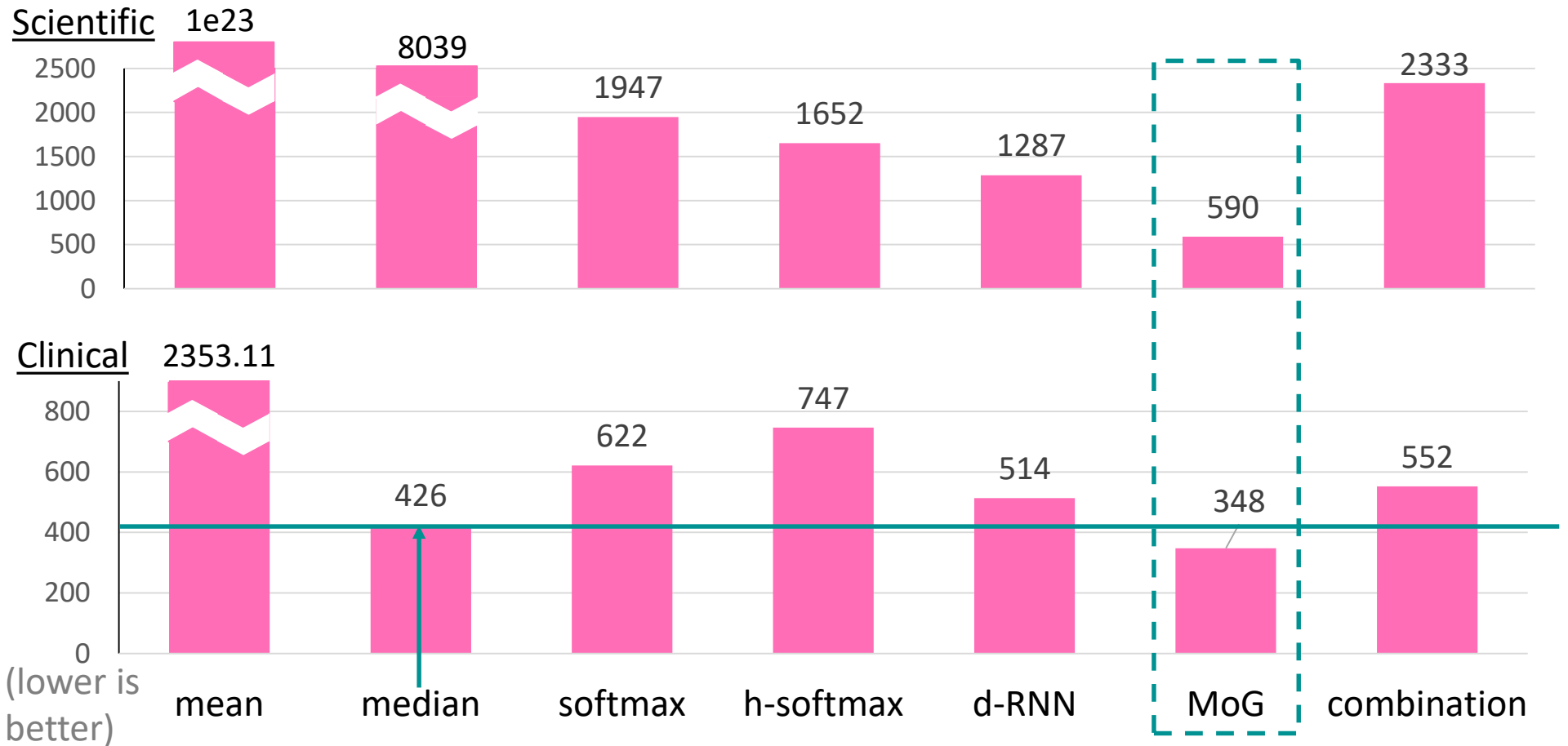
numeral → number

`2.1' 2.1

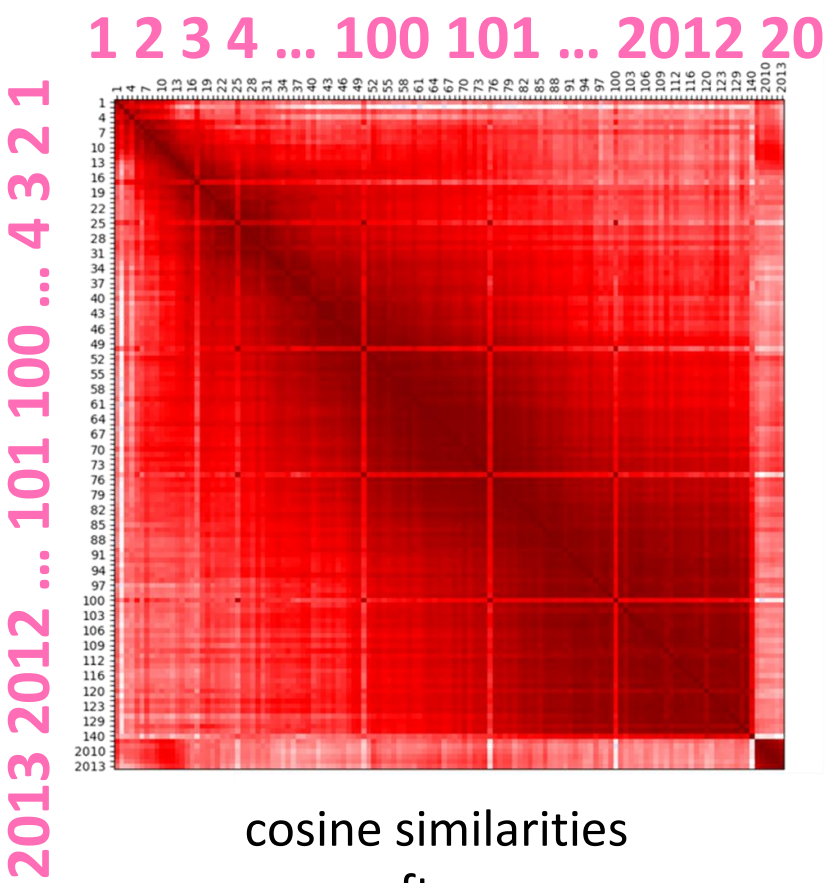
$$MAPE = \frac{prediction - target}{target} \times 100\%$$



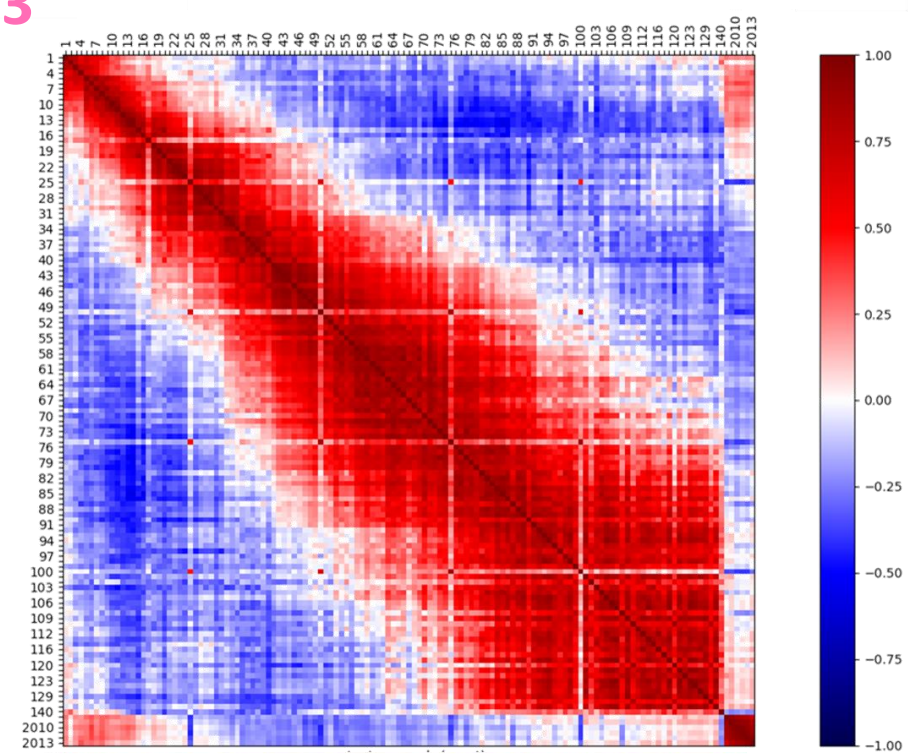
Results: Number Prediction



Softmax versus Hierarchical Softmax

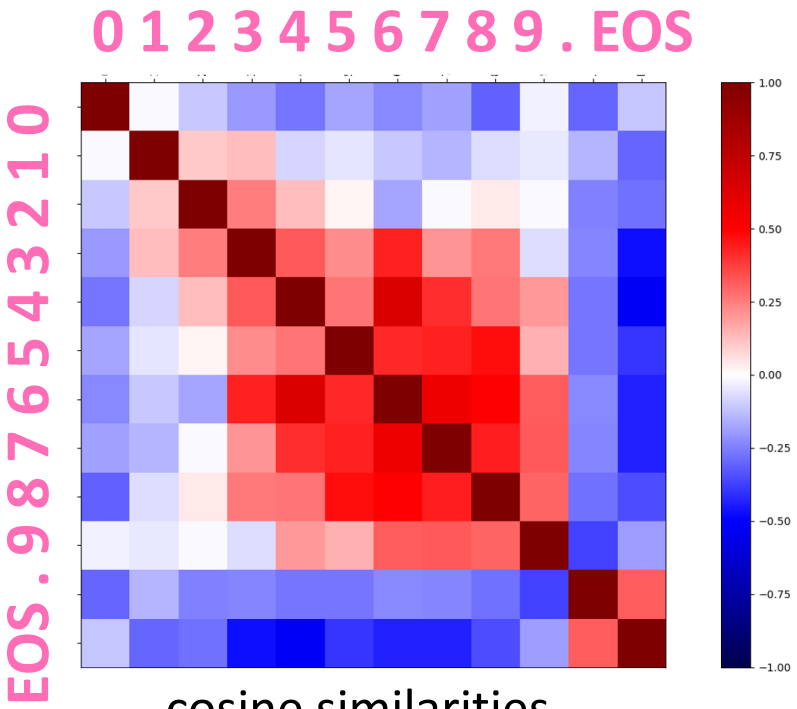


cosine similarities
softmax



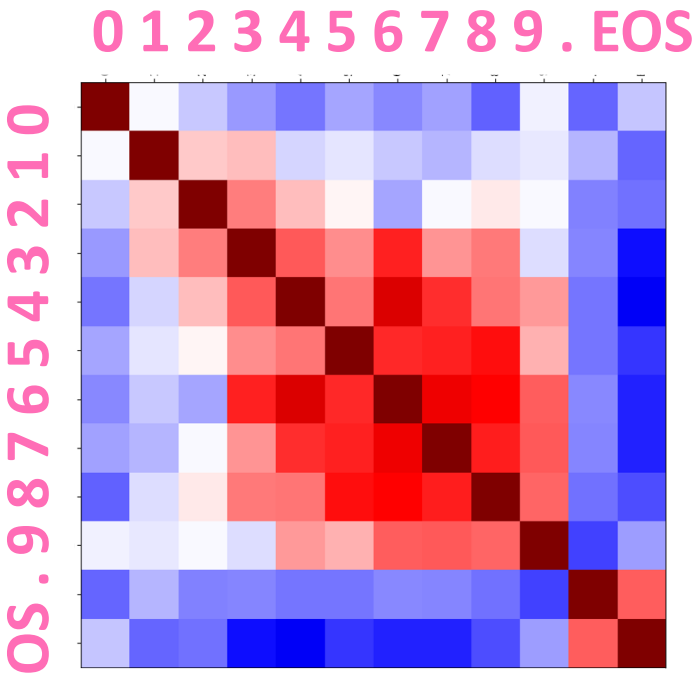
cosine similarities
h-softmax

Analysis: d-RNN and Benford's Law

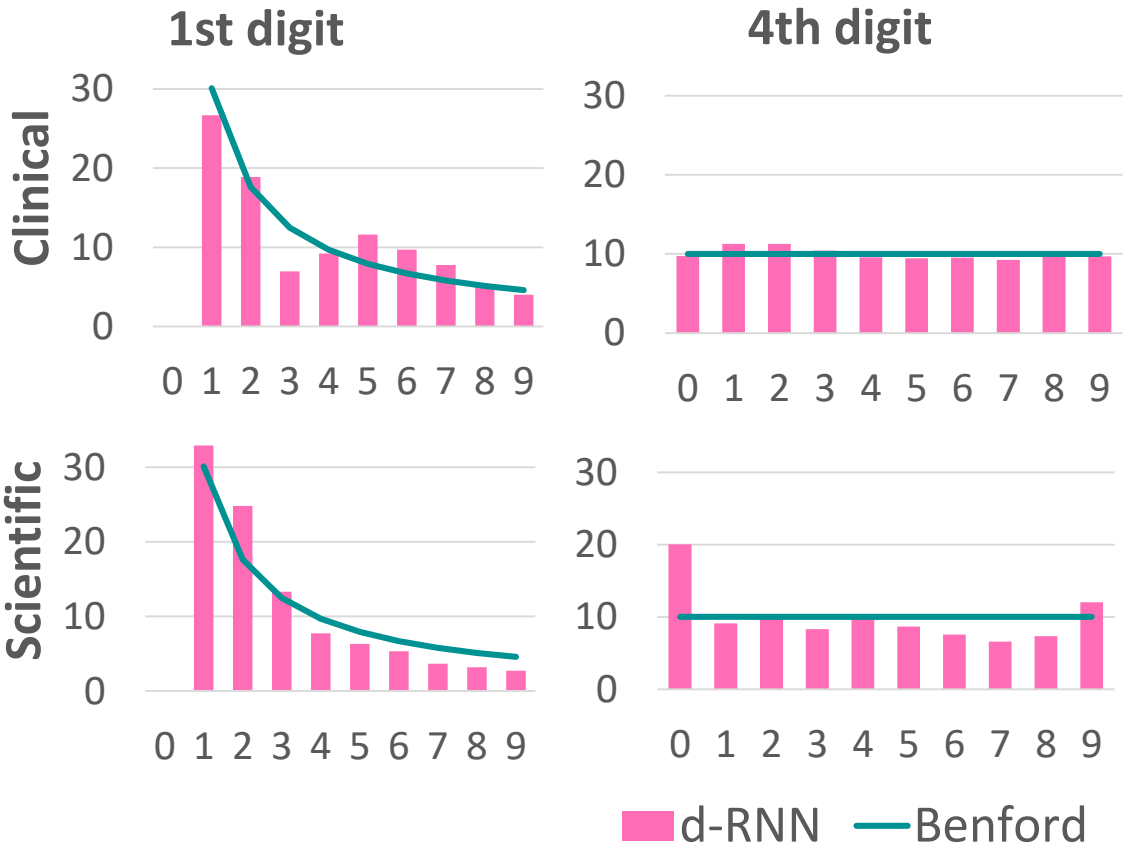


cosine similarities
d-RNN

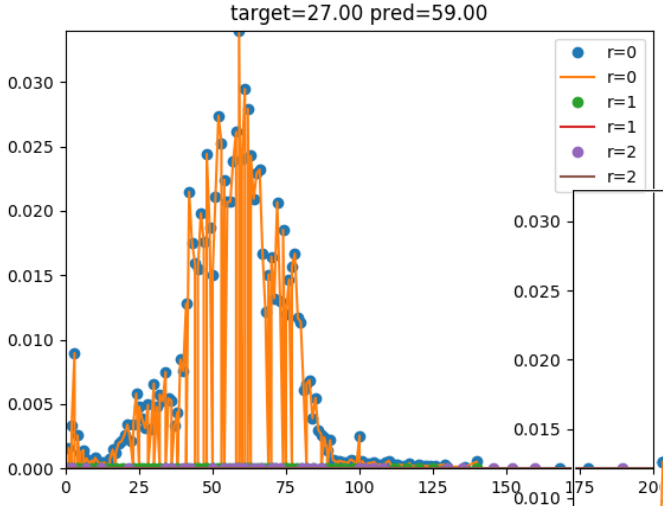
Analysis: d-RNN and Benford's Law



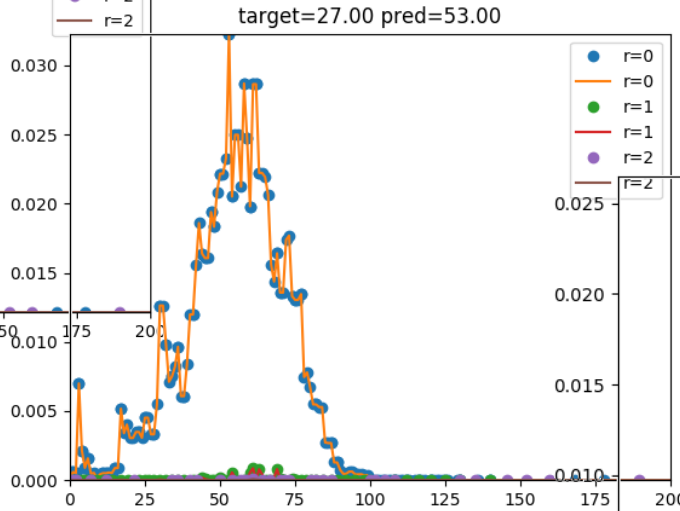
cosine similarities
d-RNN



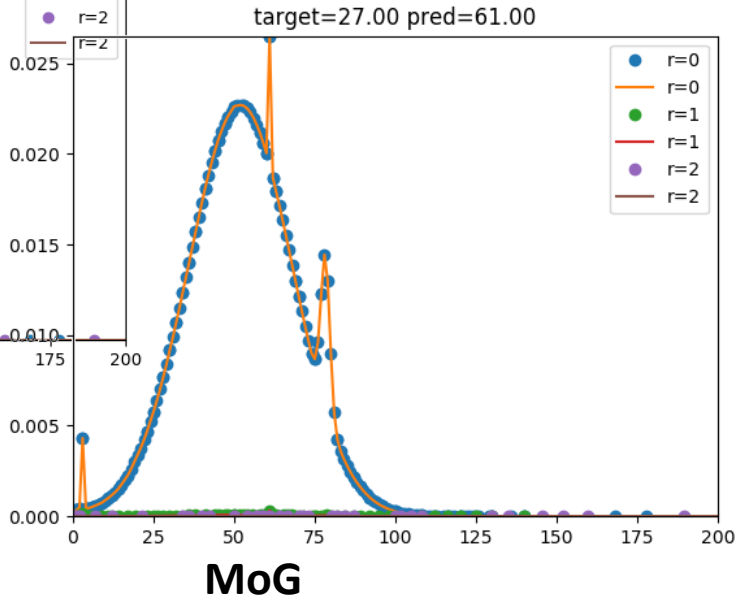
Analysis: Model Predictions



h-softmax



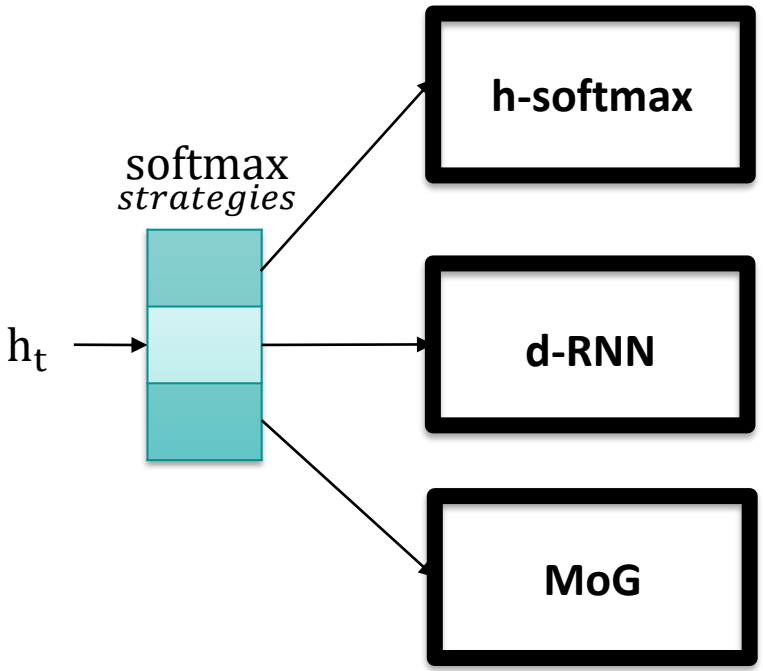
d-RNN



MoG

‘… ejective fraction : ____ % …’

Analysis: Strategy Selection



4 out of 17 segments
Enhancement > 25 %
Li et al. 2003

Small integers,
percentiles,
years

measured 32 x 31 mm
NGC 6334 stars

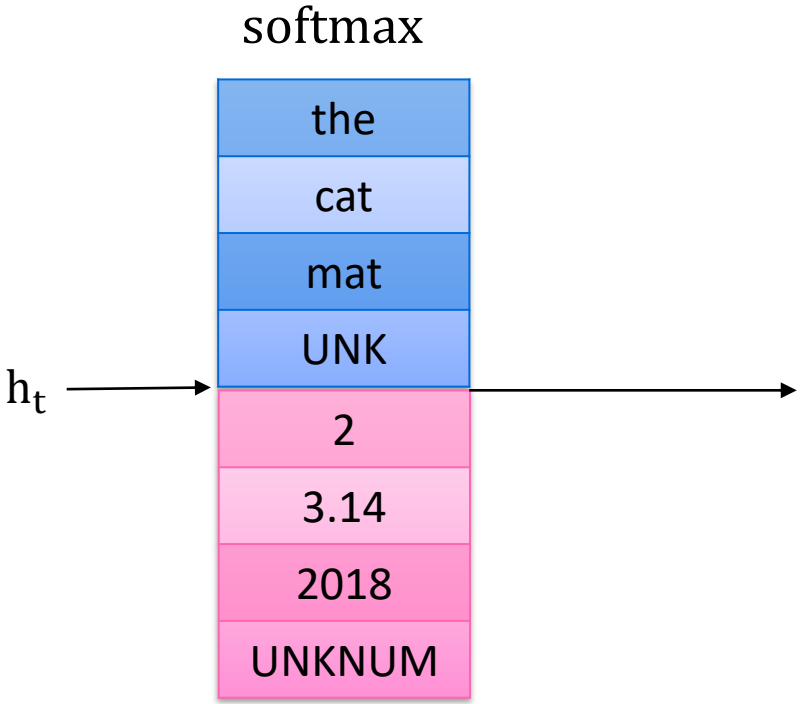
2-digit integers,
some ids

Ejective fraction: 27.00 %
Ejective fraction: 35.00 %
HIP 12961 and GL 676

reals,
some ids

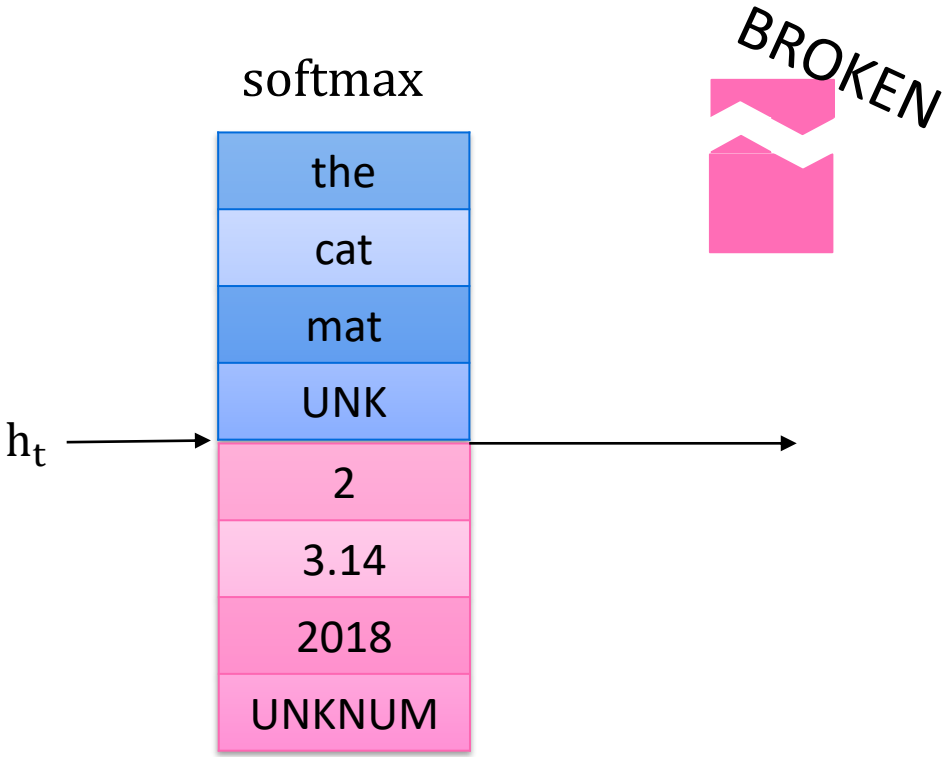
Conclusion (1)

Are existing LMs numerate?



Conclusion (1)

Are existing LMs numerate?



BROKEN

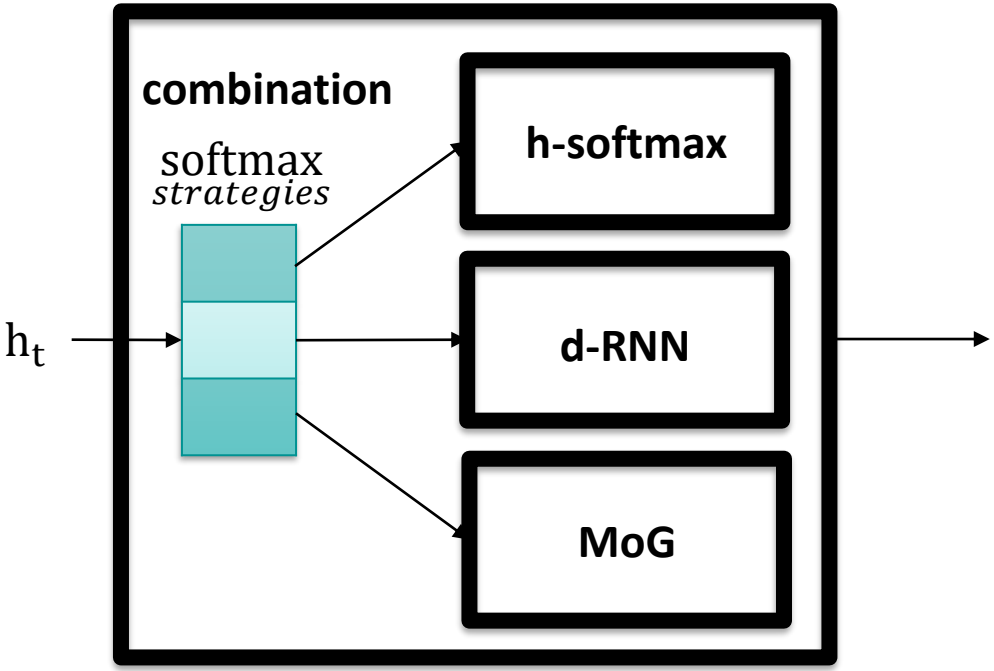
‘John’s height is ___ ’

0 2 1 25 999 3 UNKNUM 3.14 2018 50

Conclusion (2)

How to improve

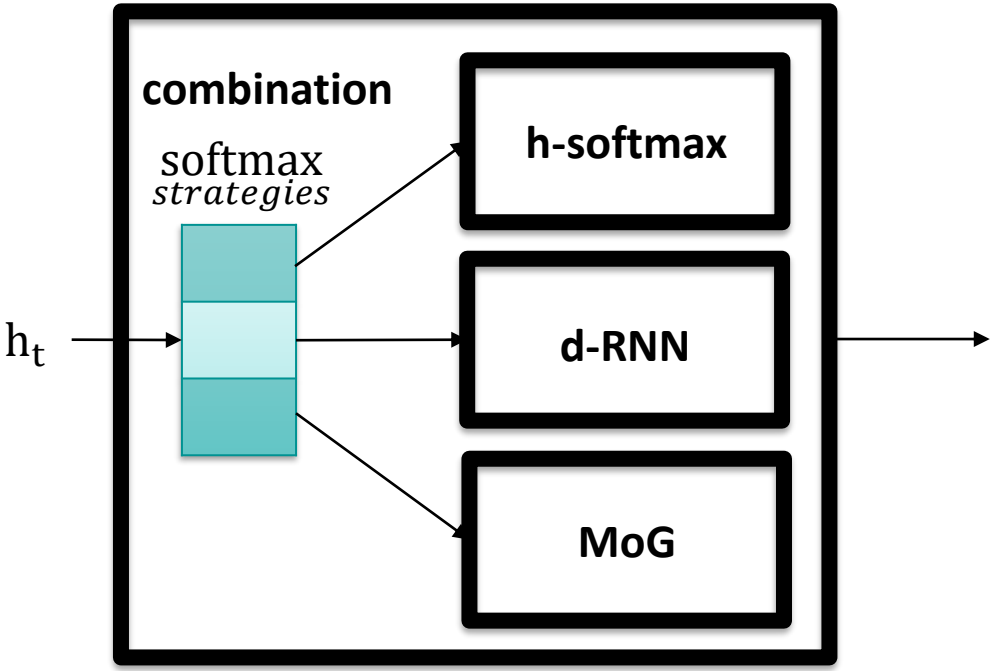
the numeracy of LMs?



Conclusion (2)

How to improve

the numeracy of LMs?



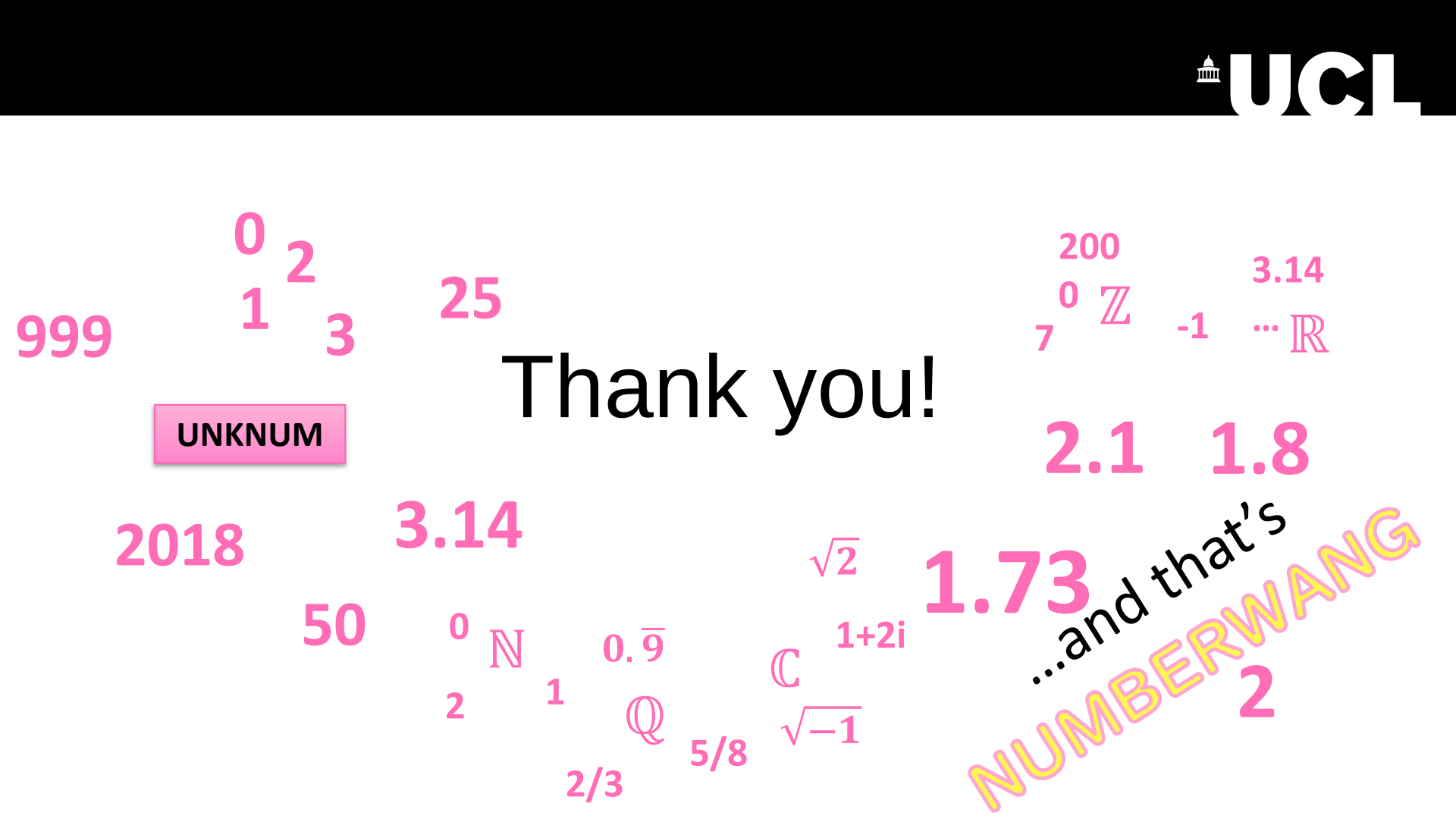
‘John’s height is ___ ’

2.1

1.8

1.73

2


0
2
1
3
25
999
200
0
$\mathbb{Z}$
7
-1
3.14
... $\mathbb{R}$
Thank you!
UNKNUM
2.1
1.8
2018
3.14
$\sqrt{2}$
1.73
...and that's
50
0
$\mathbb{N}$
$0.\overline{9}$
1+2i
$\mathbb{C}$
NUMBERWANG
2
1
$\mathbb{Q}$
$\sqrt{-1}$
5/8
2/3