Intention prediction in a dialog as a sequence prediction task

Demidovskij A.V.

Karpov N.V.

Motivation - Intent analysis

Intention *is a mental state that represents a commitment to carrying out an action or actions in the future. Intention involves mental activities such as planning and forethought.*

(Bratman, 1987)

There are 25 classes of intentions and 5 super classes

Таблица 1. Классификация интенций по Хабермасу

Информативно-воспроизводящий тип, Воспроизвести в речи наблюдаемое		Эмотивно-консолидирующий тип, Предложение собственной картины мира для кооперативного взаимодействия		Манипулятивный тип, Социальное доминирование, установление иерархии		Волюнтивно-директивный тип, Побудить адресата к действию, внести изменения в фрагмент действительности		Контрольно-реактивный тип, Выразить оценочную реакцию на ситуацию	
А	Удивление Вопрос	**Е**	Самопрезентация	**Л**	Оскорбления	**Р**	Побуждение к позитивному действию/ рекомендации	**Х**	Одобрение Похвала
Б	Выражение несогласия Сомнение	**Ж**	Привлечение внимания (рассуждения, риторические вопросы)	**М**	Устрашение, угрозы	**С**	Подстрекательство к негативному	**Ц**	Сарказм Злорадство
В	Выражение согласия Поддержка	**З**	Успокоение аудитории Утешение	**Н**	Дискредитация (подрыв авторитета)	**Т**	Обвинение	**Ч**	Критика
Г	Непринятие, отказ от коммуникации	**И**	Прогнозы, претензии на истину	**О**	Демонстрация силы (без явной угрозы)	**У**	Предупреждение о последствиях	**Ш**	Ирония
Д	Сочувствие, симпатия	**К**	Оправдание (как самооправдание, т.е. без обвинений)	**П**	Морализаторство, поучения	**Ф**	Отвод обвинений (если обвиняют)	**Щ**	Разоблачение

Таблица 2. Направленность интенций

№	Трактовка	Пример
1	"Я"/"Мы комментатор о себе и о "нас"	я не согласен я считаю мы вместе можем и т.п.
2	"Они автор/ы статьи	они правы они не правы они заблуждаются и т.п.
3	"Они другие комментаторы	они не понимают говорят глупость и т.п.
4	кто-то вообще, не "Я"/"Мы не "Они"	Россия Америка капиталисты и т.п.

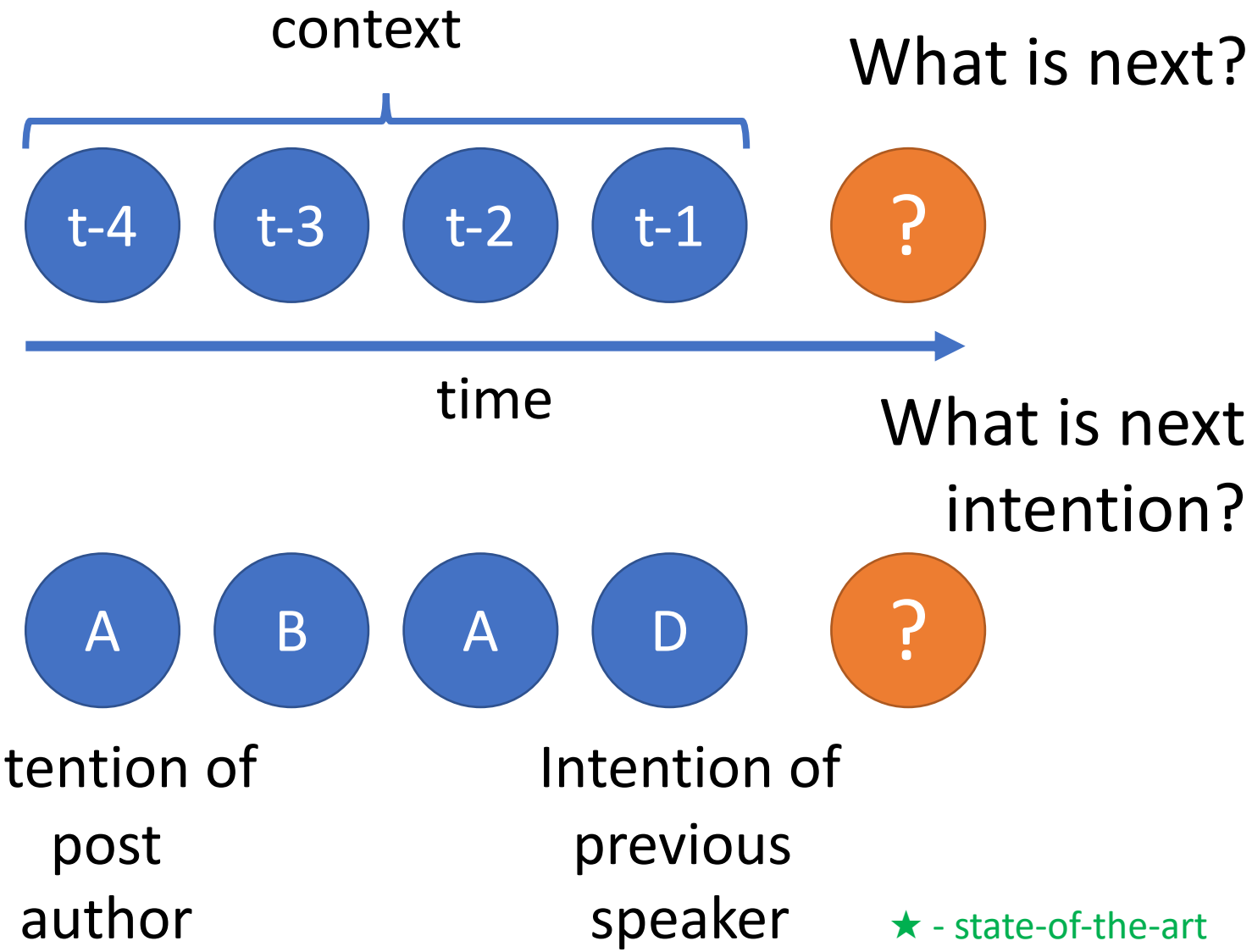
Sequence Prediction task

Implications:

- Web Page Prefetching
- Product Recommendation
- Weather Forecasting
- Opinion manipulation

Approaches:

- CPT/CPT+ (★)
- PPM, AKOM, TDAG (Markovian chains)
- DG
- LSTM



★ - state-of-the-art

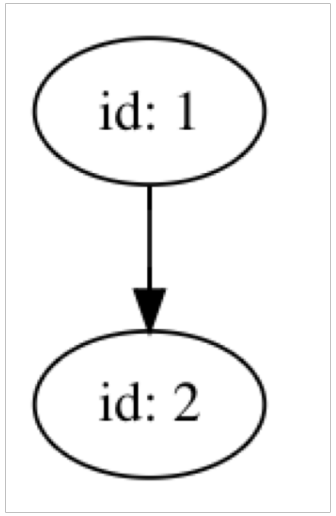
Raw Data

Format:

Graph Modeling Language

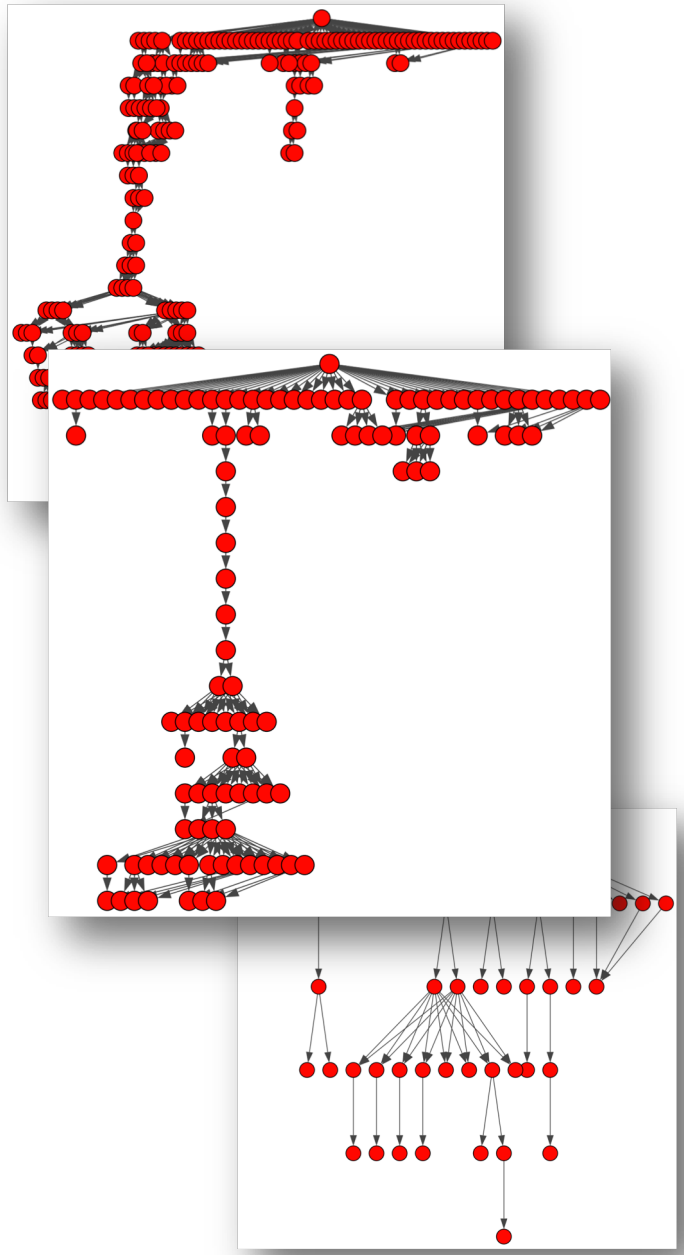
Format example:

```
graph
[
   node
   [
      id 1
   ]
   node
   [
      id 2
   ]
   edge
   [
      source 1
      target 2
   ]
]
```

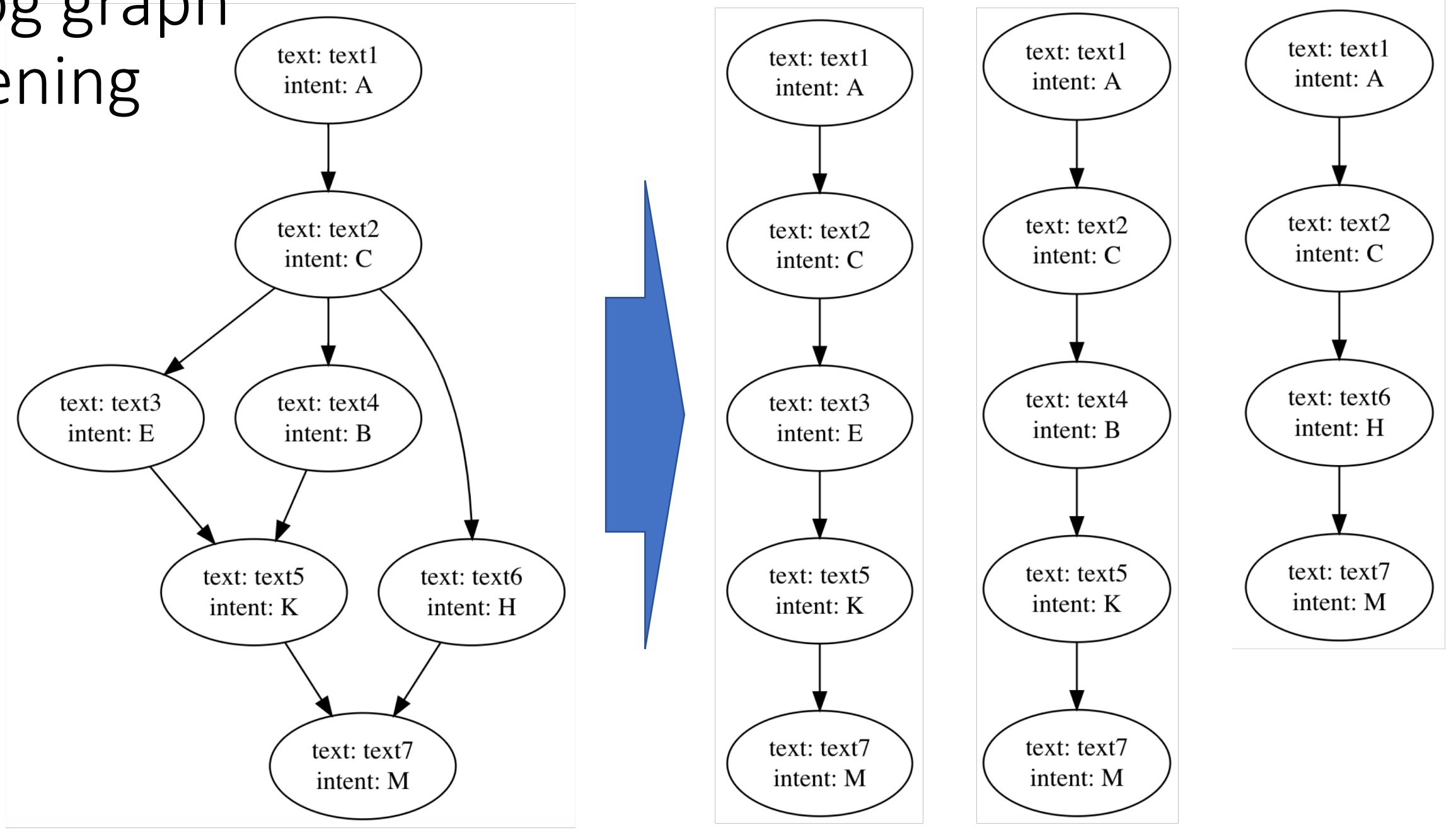


Example data

```
node
 [
  id 202505
  intent "ж"
  content "30"
  target "4"
  Label "202505"
  likes 1
  text "Тока что по Евроньюс был
        сюжет о жизни
        НЕзависимого Донецка и
        о том как он интегрируется
        через рубли в Российское
        пространство!
        Я в шоке!!))"
  date "27.05.2015 10:42"
  author "951667"
  name "202505"
 ]
```

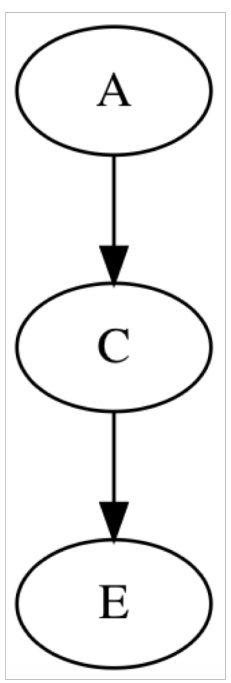


Dialog graph flattening

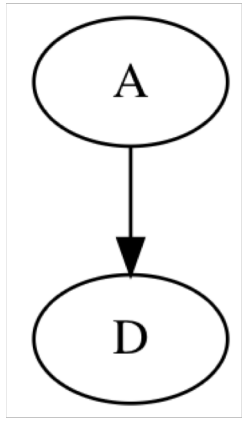


CPT/CPT+

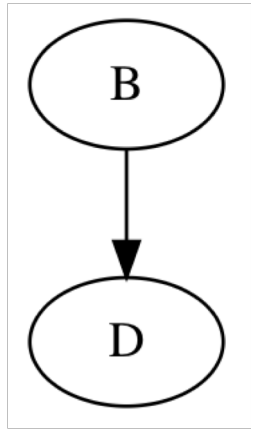
Training data



Seq. 1



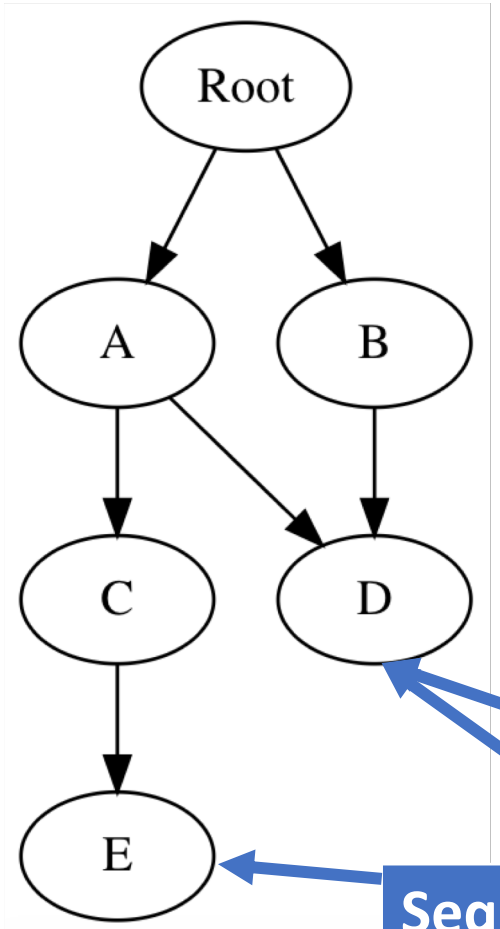
Seq. 2



Seq. 3

Model

1. Prediction Tree



3. Inverted Index

Item	Seq. 1	Seq. 2	Seq. 3
A	1	1	0
B	0	0	1
C	1	0	0
D	0	1	1
E	1	0	0

2. Lookup table

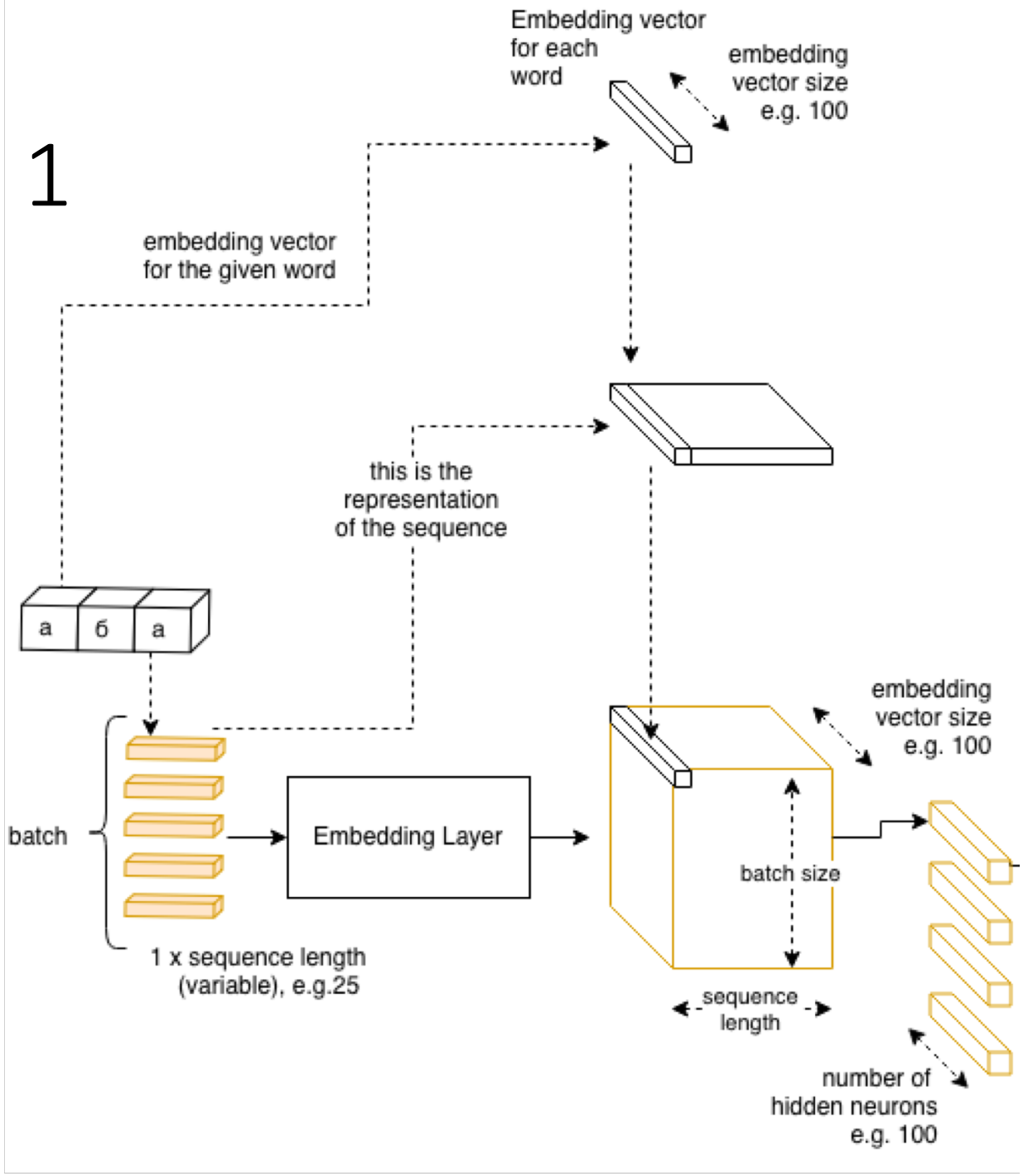
Seq. 1	Seq. 2	Seq. 3

CPT: Gueniche T., Fournier-Viger P., Tseng V.S. Compact prediction tree: A lossless model for accurate sequence prediction Springer, 2013. 177–188 с.

CPT+: Gueniche T. [и др.]. CPT+: Decreasing the time/space complexity of the Compact Prediction Tree Springer, 2015. 625–636 с.

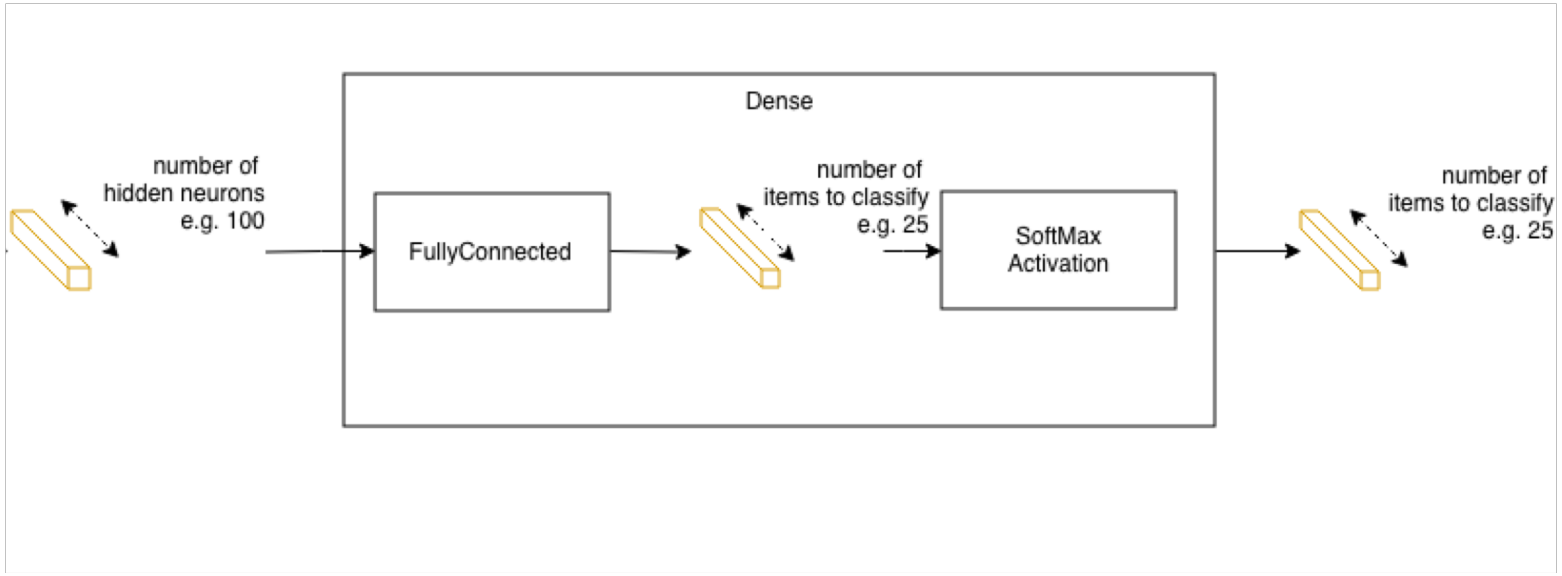
Common ANN part - 1

At the beginning of each network we put **trainable Embedding** layer to learn representations for each intention

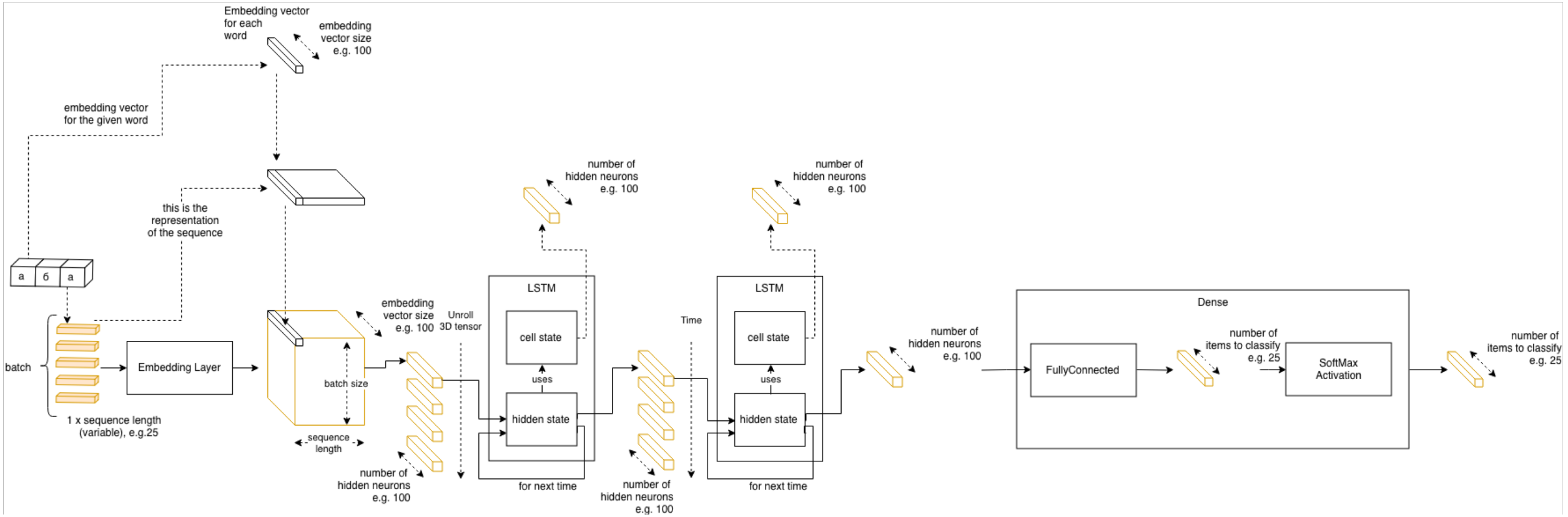


Common ANN part - 2

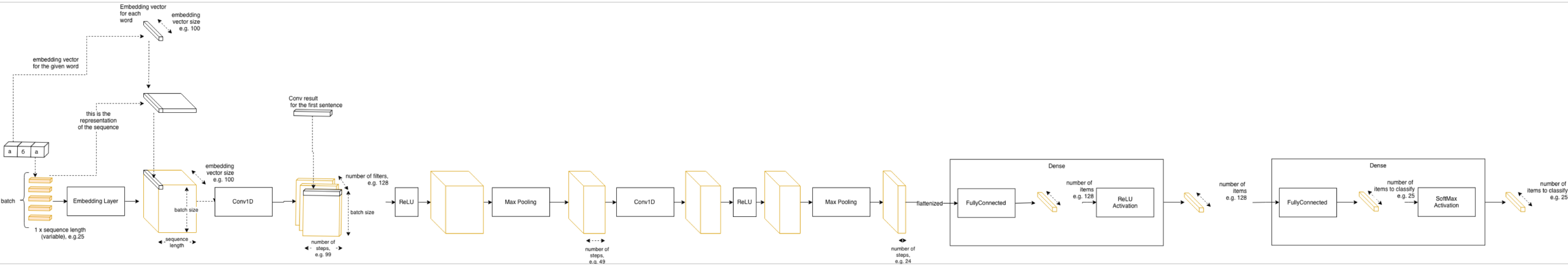
At the **end** of each network we put Fully Connected layer(s) and the SoftMax activation to get positive normalized log-predictions

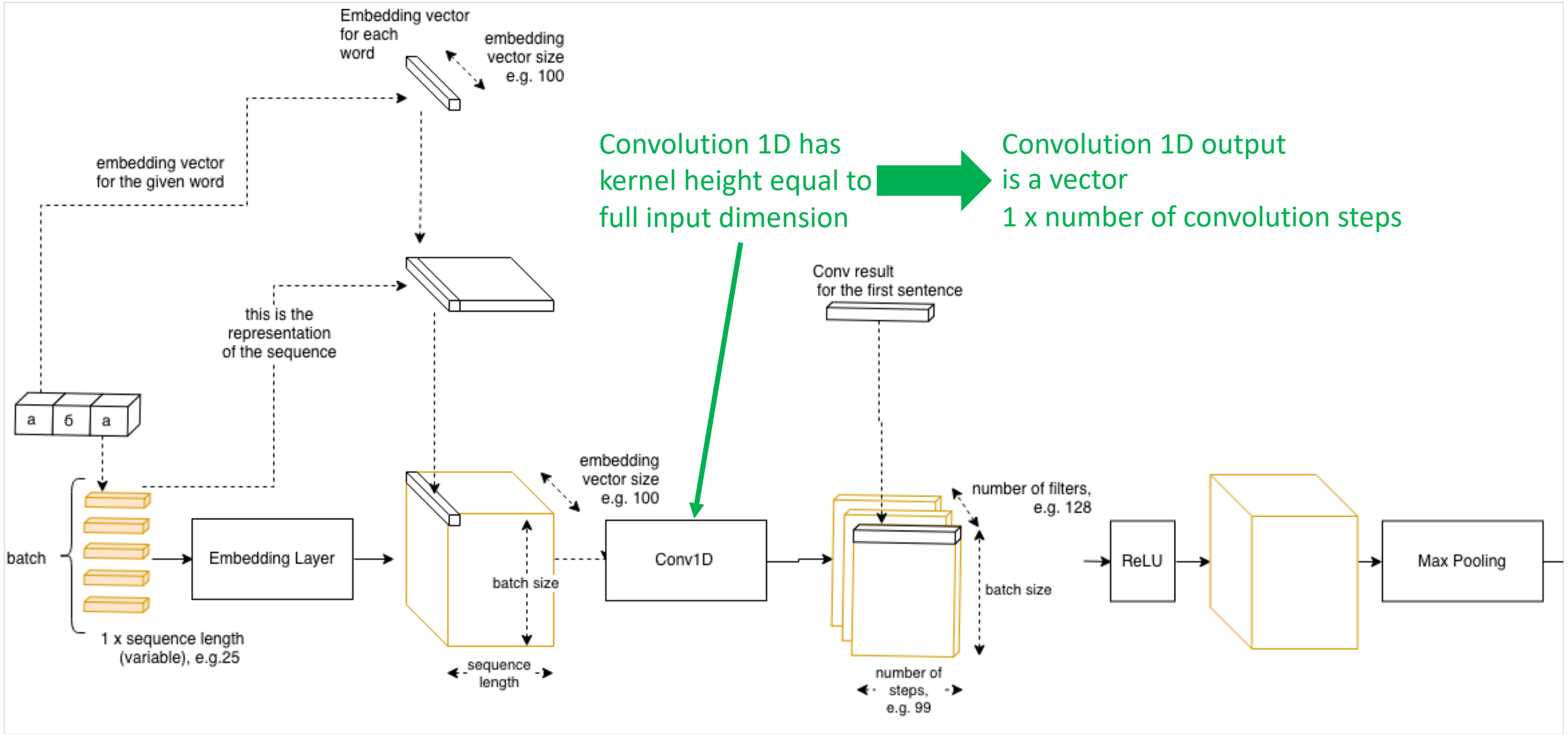


LSTM Design


Embedding vector for each word
embedding vector size e.g. 100
embedding vector for the given word
this is the representation of the sequence
а
б
а
batch
Embedding Layer
1 x sequence length (variable), e.g.25
batch size
sequence length
embedding vector size e.g. 100
Unroll 3D tensor
number of hidden neurons e.g. 100
LSTM
cell state
uses
hidden state
for next time
Time
number of hidden neurons e.g. 100
LSTM
cell state
uses
hidden state
for next time
number of hidden neurons e.g. 100
Dense
FullyConnected
number of items to classify e.g. 25
SoftMax Activation
number of items to classify e.g. 25

CNN Design


Embedding vector for each word
embedding vector size e.g. 100
embedding vector for the given word
this is the representation of the sequence
Conv result for the first sentence
а
б
а
batch
Embedding Layer
1 x sequence length (variable), e.g.25
embedding vector size e.g. 100
batch size
sequence length
Conv1D
number of filters, e.g. 128
batch size
number of steps, e.g. 99
ReLU
Max Pooling
number of steps, e.g. 49
Conv1D
ReLU
Max Pooling
number of steps, e.g. 24
flattenized
Dense
FullyConnected
number of items e.g. 128
ReLU Activation
number of items e.g. 128
Dense
FullyConnected
number of items to classify e.g. 25
SoftMax Activation
number of items to classify e.g. 25


Embedding vector
for each
word
embedding
vector size
e.g. 100
embedding vector
for the given word
this is the
representation
of the sequence
а
б
а
batch
Embedding Layer
1 x sequence length
(variable), e.g.25
batch size
sequence
length
embedding
vector size
e.g. 100
Conv1D
Conv result
for the first sentence
number of filters,
e.g. 128
batch size
number of
steps,
e.g. 99
ReLU
Max Pooling
Convolution 1D has
kernel height equal to
full input dimension
Convolution 1D output
is a vector
1 x number of convolution steps

Results

Algorithm	Accuracy (25 classes)	Accuracy (5 classes)
PPM	0,26	0,36
DG	0,23	0,38
CPT+	**0,48**	**0,66**
LSTM	0.26	0.58
CNN	**0.47**	**0.62**

Retrospective on results obtained

- Almost on par with CPT+ but still worse
- Why?
  - Considering only labels is not enough – what do we add then by the ANN?
  - Possible that the dataset was subjectively labeled

- Therefore, need to consider:
  - consider the text of each phrase itself
  - the distance between neighbor phrases
  - revalidate existing/find new dataset

Thank you for attention!