

Нейронные сети

Екатерина Черняк

Факультет компьютерных наук НИУ ВШЭ

October 9, 2018

1 Сверточные нейронные сети

- CNN для классификации предложений
- Символьные CNN для классификации предложений

Сверточные нейронные сети

Сверточные нейронные сети [англ. convolutional neural network]:

- Заимствованы из области компьютерного зрения
- Пик популярности пришелся на 2014 (до +10% аккуратности в задачах классификации), со временем были вытеснены рекуррентными нейронными сетями

Помогают справиться с двумя проблемами:

- 1 Часто входы бывают переменной длины (тексты, абзацы, предложения)
- 2 Если использовать подход, основанный на представлении предложениями окнами, то:
 - ▶ число параметров увеличивается,
 - ▶ нужно подбирать размер окна.

Слой свертки

Фильтр [англ. filter]:

$w_{1:n}$ – входная последовательность

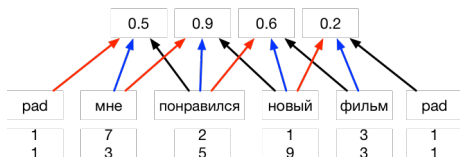
слов, E_{w_i} – эмбединг слова w_i

$x_i = \oplus(w_{i:i+k-1})$ – окно длины k

Фильтр: $p_i = g(x_i u)$

$p_i \in \mathbb{R}, x_i \in \mathbb{R}^{k \cdot d_{emb}}, u \in \mathbb{R}^{k \cdot d_{emb}}$

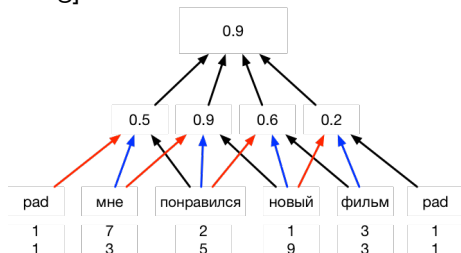
Преобразуем каждое входное окно, но пока размерность входа не уменьшается!



Слой субдискретизации (пулинга)

Субдискретизация / пулинг [англ. pooling]:

- p_i – выходные значения фильтра

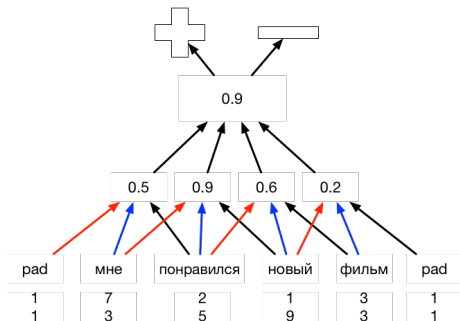


max-пулинг: $c = \max_i p_i$

- Выбираем самый важный признак из полученных на предыдущем шаге
- Можем использовать и min, и усреднение

Классификатор на основе сверточной сети

- $y \in [0, 1]$ - истинные значения
- $\hat{y} = c$ - предсказанные значения



- Для обучения сверточной сети можно использовать обычный алгоритм распространения ошибки
- Одномерные фильтры – это сильное ограничение. Что делать, если $c = 0.5$?

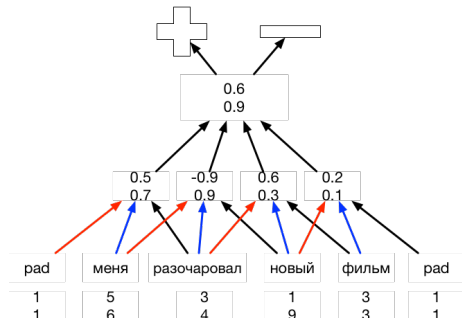
Многомерные фильтры

Применяем фильтр l раз:

$$p_i = g(x_i \cdot U + b)$$

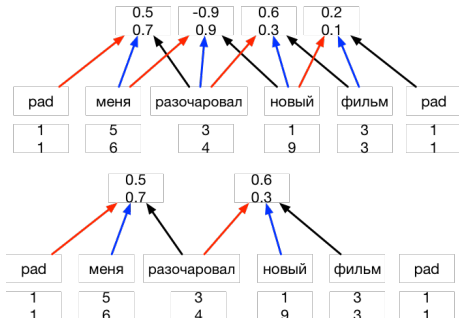
$$p_i \in \mathbb{R}^l, x_i \in \mathbb{R}^{k \cdot d_{emb}}$$

$$U \in \mathbb{R}^{k \cdot d_{emb} \times l}, b \in \mathbb{R}^l$$



Шаг окна

- Можно использовать непересекающиеся окна, чтобы уменьшить объем вычисления



Как выбирать вектора слов?

- Случайная инициализация (если нет обученных моделей word2vec, GloVe)
- word2vec, GloVe без обновления
- word2vec, GloVe с обновлением на каждой эпохе (увеличивается количество параметров!)
- Несколько каналов: копируем два входа и
 - ▶ на один подаем word2vec и не обновляем эти входы во время обучения, на второй подаем word2vec и обновляем эти входы во время обучения
 - ▶ на один вход подаем word2vec, на второй – GloVe

Как использовать pad?

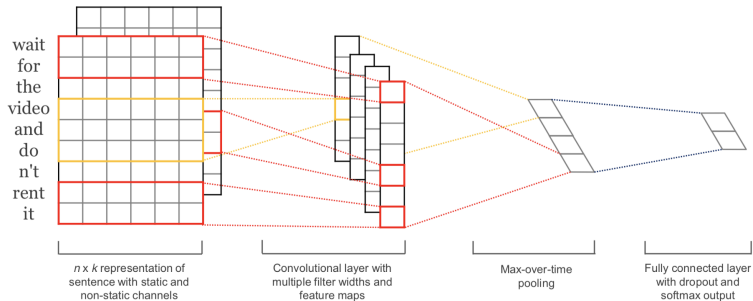
[[мое первое короткое предложение], [второе очень длинное предложение, которое никогда не заканчивается], [третье предложение]]

- 1 Неэффективный способ: одно предложение – одна эпоха
- 2 Окружить все предложения балластными символами pad и сделать их одной длины
 - ▶ Надо убедиться, что max-пулинг не выберет значения, соответствующие pad
 - ▶ Надо убрать выбросы, то есть, супер-длинные предложения, возникшие, например, из-за ошибок сегментатора

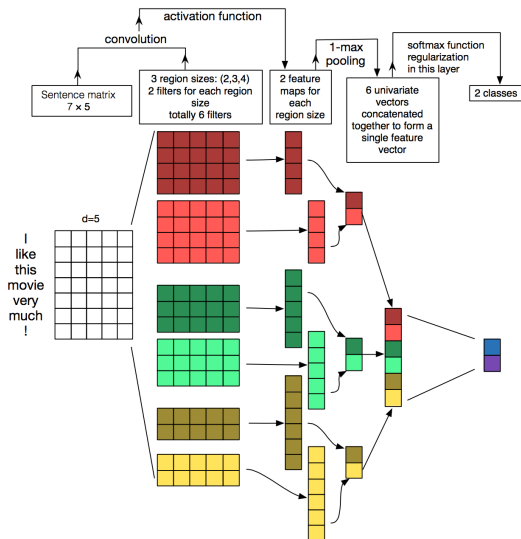
1 Сверточные нейронные сети

- CNN для классификации предложений
- Символьные CNN для классификации предложений

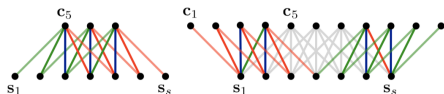
CNN для классификации предложений [Kim14]



CNN для классификации предложений [ZW15]



Dynamic Convolutional Neural Network [KGB14]



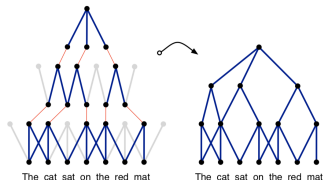
Narrow VS wide convolution:

- $m \in \mathbb{R}^m$ - weights, $s \in \mathbb{R}^s$ - input sequence
- convolution: $c_j = m^T s_{j-m+1:j}$
- narrow convolution: $s \geq m$,
 $c \in \mathbb{R}^{s-m+1}$, $j \in [m, s]$
- wide convolution: $c \in \mathbb{R}^{s+m-1}$,
 $j \in [1, s+m-1]$
- $s_i = 0, i < 1, i > s$

Dynamic Convolutional Neural Network [KGB14]

Dynamic k -max pooling:

- 1 k -max pooling over a linear sequence of values returns the subsequence of k maximum values in the sequence
- 2 the pooling parameter k can be dynamically chosen by making k a function of other aspects of the network or the input.



$$k_l = \max(k_{top}, \frac{L - l}{l} s)$$

l – the number of the current convolutional layer to which the pooling is applied

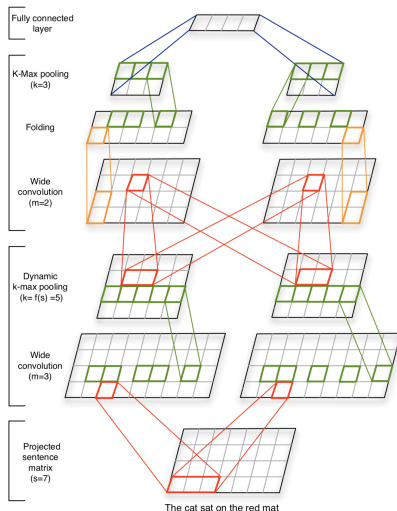
L – the total number of convolutional layers in the network

k_{top} – the fixed pooling parameter for the topmost convolutional layer

Dynamic Convolutional Neural Network [KGB14]

Folding:

After a convolutional layer and before (dynamic) k -max pooling, one just sums every two rows in a feature map component-wise.

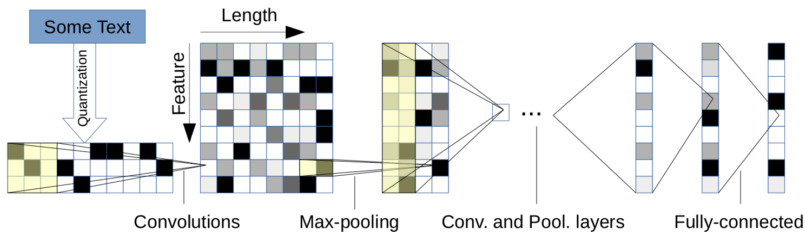


1 Сверточные нейронные сети





- CNN для классификации предложений
- Символьные CNN для классификации предложений

Символьные CNN для классификации предложений [ZZL15]

Представление текста: one-hot вектора для 70 алфавитных и неалфавитных символов



Источники I

-  Nal Kalchbrenner, Edward Grefenstette, and Phil Blunsom, *A convolutional neural network for modelling sentences*, arXiv preprint arXiv:1404.2188 (2014).
-  Yoon Kim, *Convolutional neural networks for sentence classification*, arXiv preprint arXiv:1408.5882 (2014).
-  Ye Zhang and Byron Wallace, *A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification*, arXiv preprint arXiv:1510.03820 (2015).
-  Xiang Zhang, Junbo Zhao, and Yann LeCun, *Character-level convolutional networks for text classification*, Advances in neural information processing systems, 2015, pp. 649–657.