Natural Language Processing

Яковенко Ольга

NLP tasks

Unsupervised (без учителя)	Supervised (с учителем)
▪ Topic modelling ▪ Language modelling ▪ Поиск похожих	▪ Topic recognition ▪ Sentiment recognition ▪ Spelling error detection + correction ▪ Machine translation ▪ Intent recognition ▪ Spam detection

NLP tasks

Unsupervised (без учителя)	Supervised (с учителем)
▪ Topic modelling ▪ Language modelling ▪ Поиск похожих	▪ Topic recognition ▪ Sentiment recognition ▪ Spelling error detection + correction ▪ Machine translation ▪ Intent recognition ▪ Spam detection

Sentiment recognition

Распознавание эмоциональности высказывания:

- Позитивное/негативное/нейтральное;
- Разновидности негативного (расизм, политика, уничижение соц и нац меньшинств, …)/нейтральное…

Подходы к решению

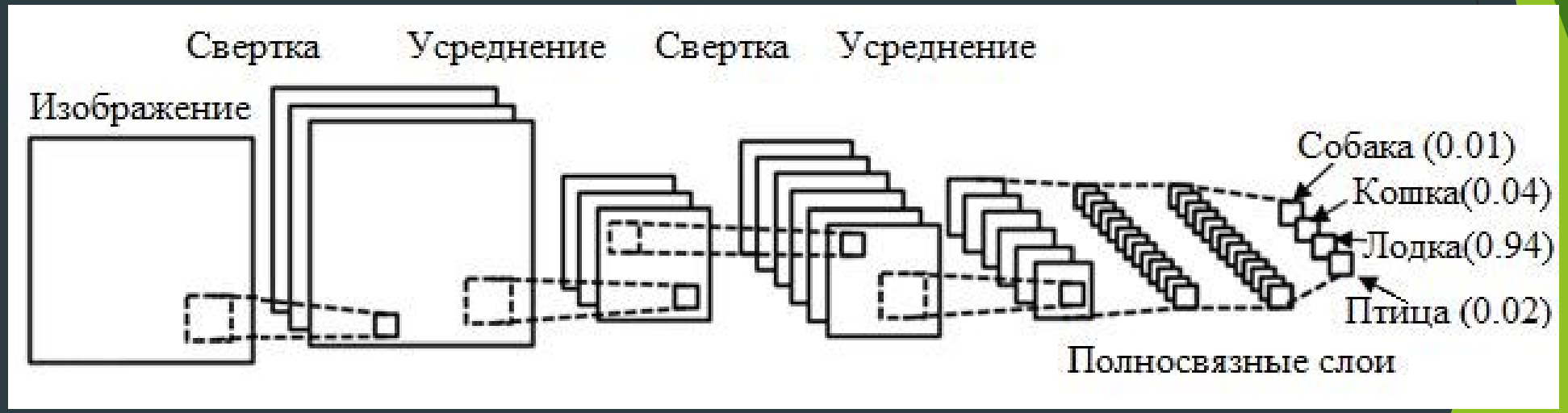
Признаковое представление одного сэмпла	Модель классификации
1D вектор, представляющий целый текст (BoW, Tf-idf, усреднённые word2vec, bag of word2vec)	Logistic Regression или Multilayer Perceptron
Матрица размера (max_n_words, feature_vector_size) – настаканные друг на друга эмбеддинги слов для представления текста (фиксированного размера)	Convolutional Neural Network
Матрица размера (n_words, feature_vector_size) – настаканные друг на друга эмбеддинги слов для представления текста (переменного размера)	Recurrent Neural Network

Задание

Как будет выглядеть матрица признаков в первом, втором и третьем случае?

Признаковое представление одного сэмпла
1D вектор, представляющий целый текст (BoW, Tf-idf, усреднённые word2vec, bag of word2vec)
Матрица размера (max_n_words, feature_vector_size) – настаканные друг на друга эмбеддинги слов для представления текста (фиксированного размера)
Матрица размера (n_words, feature_vector_size) – настаканные друг на друга эмбеддинги слов для представления текста (переменного размера)

Convolutional Neural Network (Свёрточная Нейронная Сеть)



Convolutional Neural Network (Свёрточная Нейронная Сеть)

KERAS & TENSORFLOW

Определение типа модели

Добавление слоя свёртки

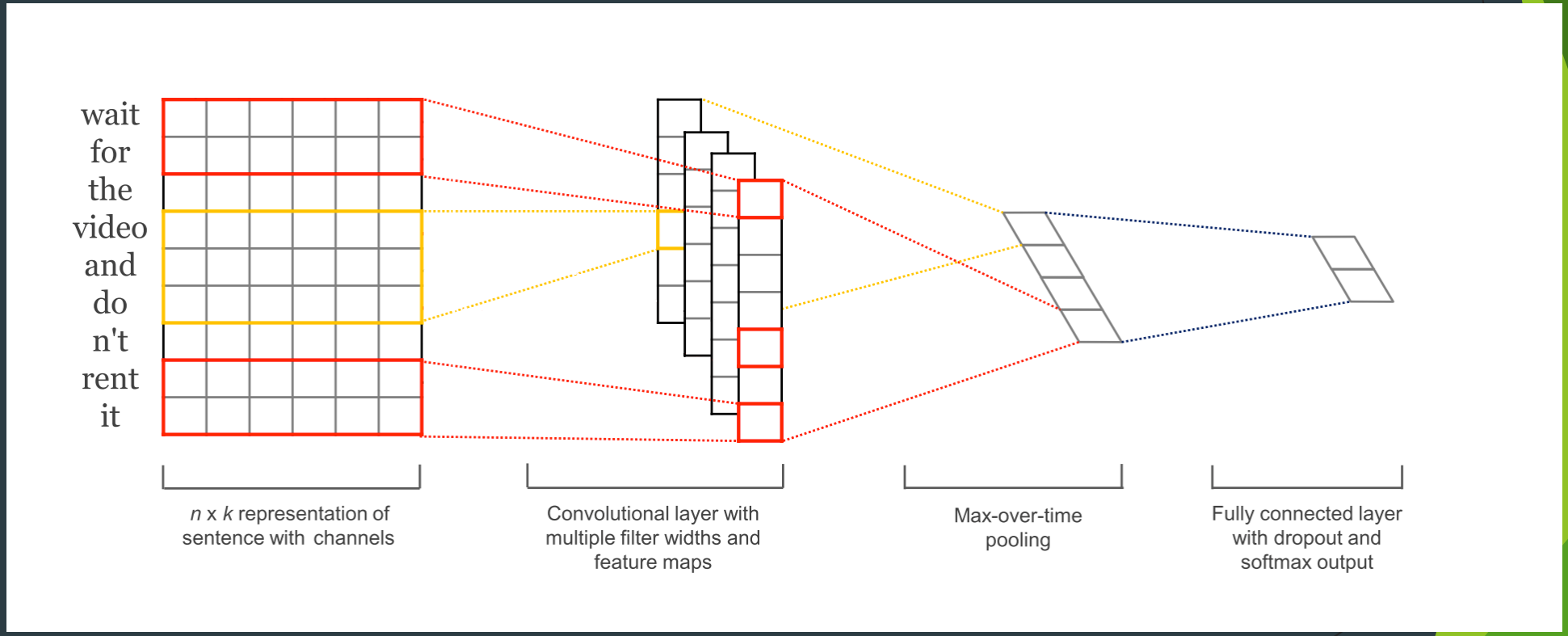
Добавление слоя подвыборки

Добавление обычного слоя с 1000 нейронами

```
model = Sequential()
model.add(Conv2D(32, kernel_size=(5, 5), strides=(1, 1),
                 activation='relu',
                 input_shape=input_shape))
model.add(MaxPooling2D(pool_size=(2, 2), strides=(2, 2)))
model.add(Conv2D(64, (5, 5), activation='relu'))
model.add(MaxPooling2D(pool_size=(2, 2)))
model.add(Flatten())
model.add(Dense(1000, activation='relu'))
model.add(Dense(num_classes, activation='softmax'))
```

https://adventuresinmachinelearning.com/keras-tutorial-cnn-11-lines/

Convolutional Neural Network (Свёрточная Нейронная Сеть)



https://www.aclweb.org/anthology/D14-1181

Практика

https://github.com/DinoTheDinosaur/russian_sentiment_edu/blob/master/notebooks/Logistic_Regression_BoW.ipynb