

Применение модели глубокого обучения sequence-to-sequence в задаче формирования аннотаций

TensorFlow/textsum

Ларюшина Юлия

Шашкин Павел

15МАГПМИ

План работы

- Изучение системы машинного обучения TensorFlow
- Установка и настройка библиотеки
- Сбор данных для обучения
- Решение распространенных проблем
- Создание tutorials для разработчиков
- Обучение модели textsum
- Создание интерфейса для взаимодействия с моделью

TensorFlow / общая информация

- Граф потока данных, представляющий вычисления
- Вершины - операции (operation)
- Ребра – тензоры (tensor)
- Вычисления реализуются в рамках сессии (session)
- Вычисления выполняются на устройствах (device) (CPU или GPU)

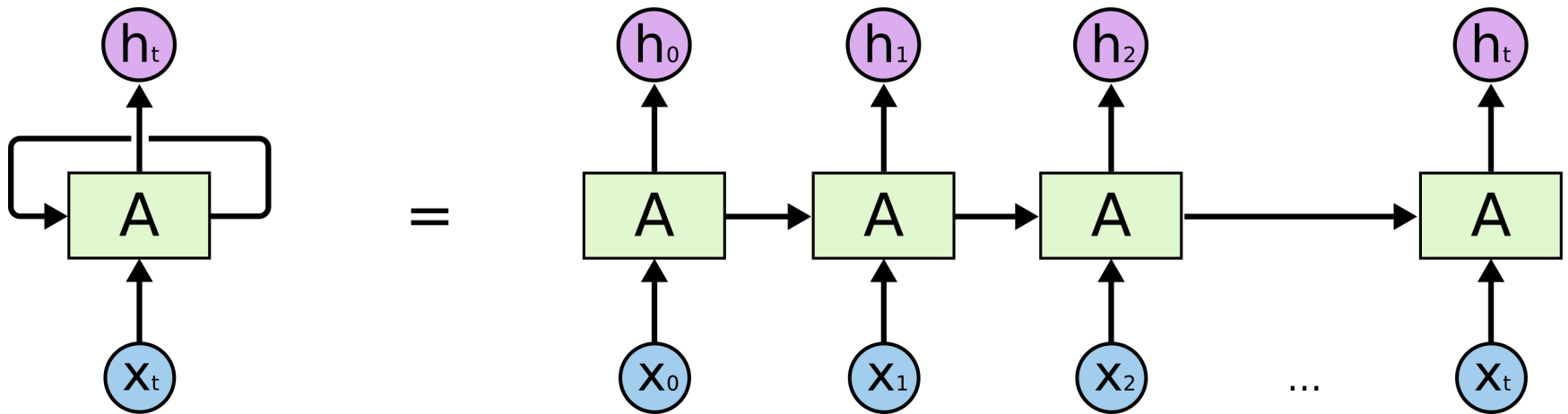
TensorFlow / общая информация

- Имеет api для Python, для R – в разработке (<https://github.com/rstudio/tensorflow>)
- TensorFlow выполняет вычисления с помощью высоко оптимизированного C++, а также поддерживает нативный API для C и C++.

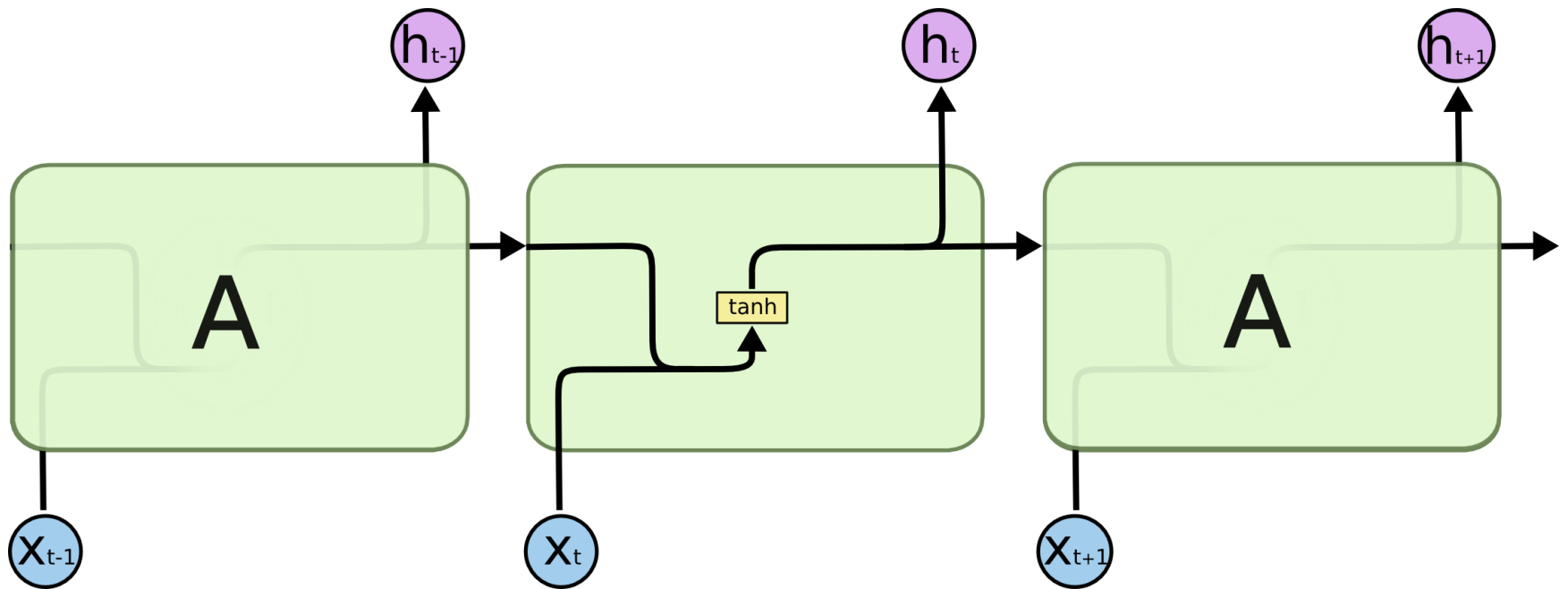
TensorFlow / простой пример

```
import tensorflow as tf
import numpy as np
matrix1 = 10 * np.random.random_sample((3, 4))
matrix2 = 10 * np.random.random_sample((4, 6))
tf_matrix1 = tf.constant(matrix1)
tf_matrix2 = tf.constant(matrix2)
tf_product = tf.matmul(tf_matrix1, tf_matrix2)
sess = tf.Session()
result = sess.run(tf_product)
sess.close()
```

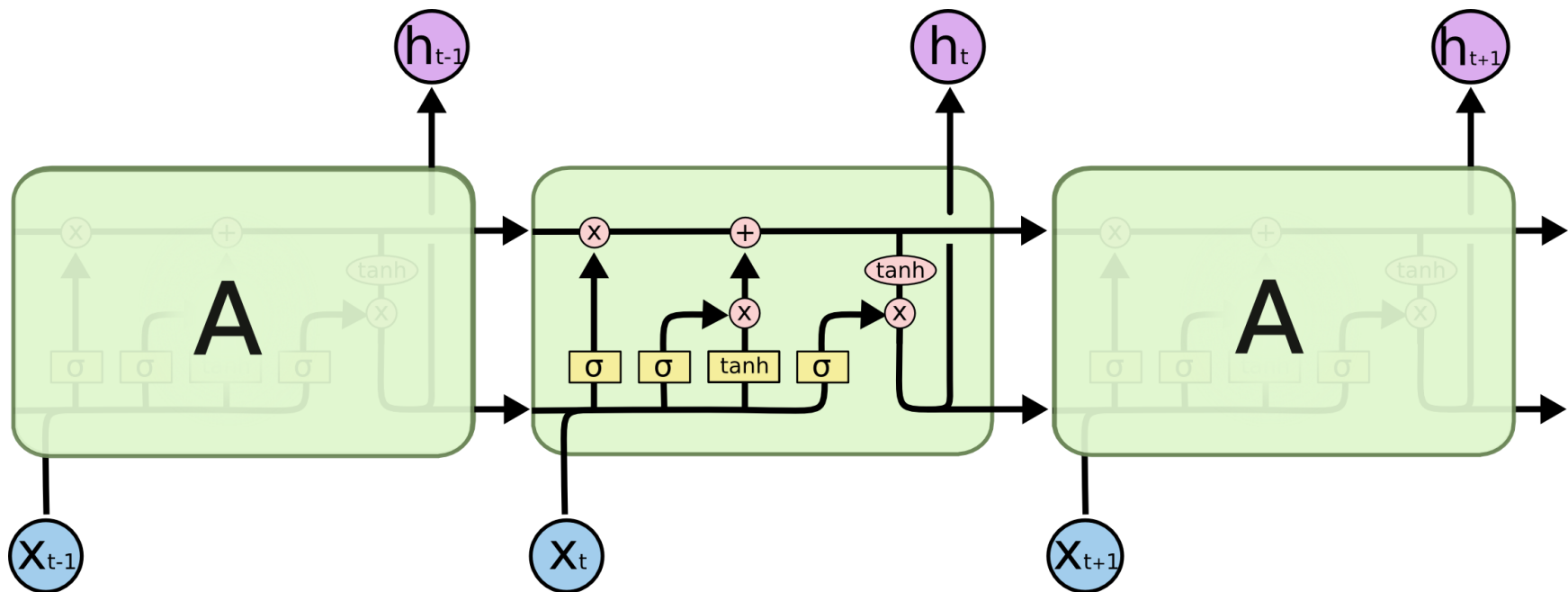
RNN



RNN traditional



LSTM



Textsum / необходимые компоненты

Подробный tutorial [здесь](#)

- TensorFlow 😊
- Bazel
- Python

TensorFlow / установка

- CPU only:
 - Python 2.7: export
TF_BINARY_URL=https://storage.googleapis.com/tensorflow/linux/cpu/tensorflow-0.11.0rc0-cp27-none-linux_x86_64.whl
 - Python 2: pip install --ignore-installed --upgrade \$TF_BINARY_URL
- Тестирование установки:

```
python
...
>>> import tensorflow as tf
>>> hello = tf.constant('Hello, TensorFlow!')
>>> sess = tf.Session()
>>> print(sess.run(hello))
Hello, TensorFlow!
>>> a = tf.constant(10)
>>> b = tf.constant(32)
>>> print(sess.run(a + b))
```

```
$ sudo apt-get install pkg-config zip g++ zlib1g-dev unzip
$ chmod +x bazel-version-installer-os.sh
$ ./bazel-version-installer-os.sh --user
$ export PATH="$PATH:$HOME/bin"
```

Подготовка окружения

В директории [workspace_sample](#) содержится пример workspace для textsum.

Необходимо воссоздать его структуру и добавить пустой WORKSPACE файл.

Имеем:

- оригинальный textsum [отсюда](#)
 - папку с "игрушечными" данными (training/data, testing/data, validation/data), пример реальных выходных данных (text_data) и словарь (vocab)
 - WORKSPACE файл (необходим для Bazel)
- Текстовый файл text_data содержит пример входных

"Игрушечный" пример

