

Анализ неструктурированных данных
Самостоятельная работа
Максимальная оценка: 10 баллов

Имя и фамилия: _____
Группа: _____

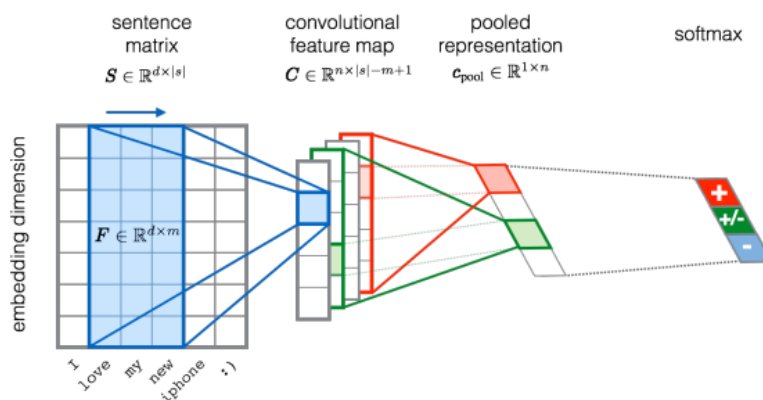
Ответьте на вопросы ниже. Можно использовать компьютер для вычислений и для просмотра лекций и учебников. Запишите решения, выделите ответы и прокомментируйте каждый шаг решения.

Каждый вопрос оценивается в 2 балла.

1. Определите

- A размерность эмбединга
- B размер окна
- C количество фильтров
- D размерность k -max pooling'a?

на приведенной ниже схеме сверточной нейронной сети.



Ответ: $d = 4$, $m = 3$, $s = 5$, $k = 1$

2. Какой тип связей между словами (длинные или короткие) лучше запоминают сверточные нейронные сети?

Ответ: короткие

3. Допустим, что мы решаем задачу тематической классификации по аннотации статьи. Динамический k -max pooling поможет:

- A определить, что слова в начале аннотации более значимы, чем в конце аннотации (будут выбраны разные k для начала и конца аннотации);
- B придать больший вес именованным сущностям (будут выбраны k помогающие определить границы именованных сущностей);
- C определить тематические регионы (k поможет выбрать нужное количество тем);
- D дообучить эмбединги (k определит необходимое количество отрицательных контекстов).

Ответ: A

4. Может ли быть решена без учителя задача определения близких по смыслу предложений? Если да, каким методом?

Ответ: WMD, $\cos(d2v)$

5. Используются ли модели класса энкодер-декодер для вычисления эмбедингов предложений? Если да, назовите хотя бы одну модель. Если нет, почему?

Ответ: skip-thoughts, GUS