

1. Ухудшилось или улучшилось бы качество результата, если бы модель предсказывала слова, а не символы?

Результат бы улучшился. Например, чтобы предсказать до третьего слова от начала предложения, RNN на уровне слов должен сделать два предсказания, чтобы получить его, тогда как RNN на уровне символов должен сделать предсказания для числа раз, равного числу символов перед вторым пробелом. Чем больше прогнозов должна сделать RNN, тем более подвержен ошибкам результат.

2. Что такое блэндинг?

Простейшая схема стекинга (ансамблирование алгоритмов) — блэндинг: обучающую выборку делят на две части. На первой обучают базовые алгоритмы. Затем получают их ответы на второй части и на тестовой выборке. Ответ каждого алгоритма можно рассматривать как новый признак (т.н. «метапризнак»). На метапризнаках второй части обучения настраивают метаалгоритм. Затем запускают его на метапризнаках теста и получают ответ.

3. Почему выбрано именно такое значение прореживания?

Обычно значение выбирается из промежутка 0.2-0.5, чтобы не исключить слишком мало/много нейронов. Исходя из этого и из того, что сеть справлялась со своей задачей, было выбрано и оставлено такое значение.