

1. Чем можно обосновать заикленность генерируемых фраз?

Возможна ситуация, когда после одного символа с большей вероятностью будет определенный символ, после которого в свою очередь тоже есть наиболее вероятный символ, и на каком то символе следующий наиболее вероятный будет из тех, который мы уже встречали, тем самым получается кольцо, то есть получаем заикленность. Например, у каждого символа есть следующий наиболее вероятный, у буквы А самый вероятный Б, у Б символ В, а у В символ А, вот и получается АБВАБВАБВ.

2. В чем заключается смысл ансамблирования комитетом большинства?

Комитет большинства:

$$a(x) = \text{mode}(b_1(x), \dots, b_n(x)).$$

где mode – мода (значение, которое встречается чаще других среди аргументов функции). Если рассмотреть задачу классификации с двумя классами $\{0, 1\}$ и три алгоритма, каждый из которых ошибается с вероятностью p , то в предположении, что их ответы – независимые случайные величины, получаем, что комитет большинства этих трёх алгоритмов ошибается с вероятностью $pp(3-2p)$. Из формулы

$$a(x) = \frac{1}{n}(b_1(x) + \dots + b_n(x))$$

видно, что это выражение может быть существенно меньше p (при $p=0.1$ почти в два раза), т.е. использование такого ансамбля уменьшает ошибку базовых алгоритмов.

3. Почему выбрано именно такое значение прореживания?

В основном между скрытыми слоями dropout используется в диапазоне 0.2 – 0.5. В данной работе значение было выбрано равным 0.2 во избежание переобучения, но также, чтобы выключить немного нейронов.