

1. В чем плюсы и минусы того, если модель предсказывала слова, а не символы?

Плюсы: были бы нормальные слова, а не просто несвязанные последовательности символов

Минусы: потребовалось бы больше памяти для генерации слов, при этом количество слов безгранично, поэтому нужно было бы увеличить количество выходов, чтобы можно было выдавать все возможные слова.

2. В чем заключается прием ECOC (error-correcting output coding)? Для чего используется?

ECOC – метод разложения многоклассовой задачи классификации на множество задач двоичной классификации, а затем объединение результатов подзадач в решение исходной задачи. Данный метод может быть использован для решения проблем текстовой классификации.