

РЕШЕНИЕ ЗАДАЧИ ОБРАТНОЙ НОРМАЛИЗАЦИИ ТЕКСТА С ПОМОЩЬЮ НЕЙРОННЫХ СЕТЕЙ

Чернышёв Александр Владиславович

Студент

Факультет ВМК МГУ имени М. В. Ломоносова, Москва, Россия

E-mail: alexander.v.chernyshev@mail.ru

Научный руководитель — Дьяконов Александр Геннадьевич

На сегодняшний день системы распознавания и синтеза речи находят большое применение в задачах голосового поиска, голосового управления и многих других, поэтому важной задачей является улучшение работы таких систем.

В системах синтеза речи приходится работать с написанными текстами, в которых есть небуквенные символы (цифры, знаки валюты и др.), из-за чего системе нужно уметь произносить, например, числительные с правильным спряжением. Для решения данной проблемы используют нормализацию текста, то есть перевод письменного предложения в его произнесённый вариант, в котором все небуквенные символы заменены словами так, как если бы эти символы были произнесены.

Аналогично, в системах распознавания речи возникает проблема с печатью распознанного текста, поэтому следует провести процедуру обратной нормализации текста — перевести текст из произнесенного варианта в письменный.

Самые простые подходы к решению задачи обратной нормализации текста используют статистические языковые модели [1] или конечные автоматы с выходом [2-3]. Такие подходы отличаются быстротой работы и не очень высоким качеством решения задачи. В более сложных подходах используются нейронные сети [3-5]. Такие подходы работают медленнее, но позволяют добиться лучшего результата.

В работе рассматриваются различные подходы к решению задачи обратной нормализации текста, основанные на нейронных сетях. Выдвигается предположение о том, что информация о типах токенов (например, числительные, знаки препинания и др.) может улучшить качество моделей. Такая информация может быть представлена как в виде непосредственно типа токена, так и в виде типового эмбединга токена, полученного с помощью энкодера трансформера.

В качестве базовых моделей использовались рекуррентная нейронная сеть и трансформер. Далее исследовались различные способы внедрения информации о типах токенов при подаче входных по-

следовательностей в трансформер: маркировка каждого токена его типом (модели №2 и №4), объединение входных токенов в группы по типу (модели №3 и №5), конкатенация эмбеддингов токенов с их типовыми эмбеддингами (модели №1 и №6), преобразование только токенов заданных типов (модели №4, №5 и №6).

Все предложенные модели превзошли базовые по метрике WER (таблица 1, чем меньше WER — тем лучше), из чего можно сделать вывод, что использование типовой информации действительно улучшает качество работы моделей.

Модель	RNN	Transformer	№1	№2	№3	№4	№5	№6
WER, %	20.8	13.2	10.2	11.3	11.5	5.5	5.4	7.0

Таблица 1: Результаты тестирования моделей.

Результаты экспериментов показывают, что между способами разметки типов токенов при подаче в трансформер нет значительной разницы. Основной прирост качества достигается за счет преобразования токенов только заданных типов (модели №№4–6).

Литература

1. Yun-Cheng Ju. A Language-Modeling Approach to Inverse Text Normalization and Data Cleanup for Multimodal Voice Search Applications. — 2008
2. Yang Zhang. NeMo Inverse Text Normalization: From Development To Production. — 2021
3. Monica Sunkara. Neural Inverse Text Normalization. — 2021
4. Ernest Pusateri. A Mostly Data-driven Approach to Inverse Text Normalization. — 2017
5. Tuan Manh Lai. A Unified Transformer-based Framework for Duplex Text Normalization. — 2021