

Шамало В.В.

магистрант кафедры информатики и вычислительной техники

Кубанский государственный технологический университет

Россия, г. Краснодар

РАСПОЗНАВАНИЕ ЭМОЦИЙ И НАМЕРЕНИЙ ПОЛЬЗОВАТЕЛЯ В ДИАЛоговых СИСТЕМАХ: ОБЗОР МЕТОДОВ

Аннотация: для более точного ответа человеку, система должна понять не только смысл реплики, но и настроение говорящего и то, чего он хочет. В данной статье сведены результаты по двум задачам: распознаванию эмоций и определению намерений. Здесь разобраны способы разметки эмоций, основные корпуса, три поколения архитектур и то, как система отклоняет запросы, которые не относятся к её теме. Качество в обеих задачах зависит от того, какой объём разговора модель видит, а не от числа её параметров, как принято считать.

Ключевые слова: диалоговые системы, распознавание эмоций, классификация намерений, обработка естественного языка, контекст беседы, короткие тексты.

Shamalo V.V.

Master's Student

Kuban State Technological University

Russia, Krasnodar

RECOGNITION OF USER EMOTIONS AND INTENTS IN DIALOGUE SYSTEMS: A REVIEW OF METHODS

Abstract: generating an appropriate response requires a dialogue system to account for the user's emotional state and the goal pursued, along with parsing the meaning of the utterance. The review brings together results on emotion recognition and intent detection. Ways of describing an emotional state, reference corpora, three generations of architectures and the procedure for

rejecting out-of-scope requests are considered. The quality of both tasks is shown to depend on the amount of conversational context taken into account rather than on the number of model parameters.

Keywords: dialogue systems, emotion recognition, intent classification, natural language processing, conversational context, short texts.

Введение

Одни и те же слова человека могут нести разный смысл и требовать от системы разных реакций — всё зависит от того, с каким настроением их написали и чего от системы хотят. Разобрать предложение и вытащить из него смысл мало: без этих двух вещей ответ выйдет неточным.

Обе задачи обычно решают отдельно друг от друга, хотя в реальном разговоре они связаны. В данной статье сведено то, что наработано по каждой задаче, и показано, где у них есть общие границы: как размечать данные, сколько контекста учитывать и что делать с совсем короткими репликами человека.

Методы и исследования

Применительно к данной задаче брали работы трёх типов: те, что задают способ описывать эмоцию, те, что вводят открытые корпуса, ставшие стандартом сравнения, и те, что предлагают какие-либо архитектуры. Предпочтение всегда отдавали публикациям с воспроизводимыми замерами. Материал выстроен пошагово, от описания эмоций и выбора базы для обучения до моделирования и работы с намерениями.

Результаты

Чем описывать эмоцию. Здесь применяются два подхода. Первый основан на представлении об ограниченном наборе базовых эмоций: Экман обосновывал существование небольшого числа состояний, различаемых независимо от культурной принадлежности наблюдателя [1]. Второй описывает эмоцию положением в непрерывном пространстве

признаков, задаваемом полярными парами и интенсивностью [2]. Дискретные категории упрощают разметку, но огрубляют оттенки, тогда как непрерывные шкалы точнее и хуже согласуются между разметчиками.

На чём учить. Корпус IEMOCAP содержит записи диалогов с разметкой по категориям и по непрерывным шкалам [3]; его ограничения связаны с постановочным характером материала и участием двух спикеров. Корпус MELD построен на материале телесериала и включает многосторонние беседы, однако распределение классов в нём неравномерно: доля нейтральных реплик преобладает [4]. Проблема в том, что дисбаланс переносится на обучаемые модели.

Чем моделировать. Ранние решения классифицировали каждую реплику независимо, что ограничивало точность там, где значение определяется предшествующим ходом. В модели DialogueRNN введено отдельное представление состояния для каждого говорящего наряду с общим состоянием беседы [5], а в модели DialogueGCN диалог представлен графом, рёбра которого отражают временную близость реплик и принадлежность одному участнику [6]. Предобученные двунаправленные трансформеры повысили качество представлений за счёт учёта контекста с обеих сторон от слова [7].

Намерения и чужие запросы. Намерение определяют как обычную классификацию по списку интенгов, и главная трудность не в ней: пользователь регулярно спрашивает то, чего в списке нет. Корпус CLINC150 разработан для оценки такой способности: наряду с полутора сотнями интенгов он содержит запросы вне области, которые система обязана отклонить [8]. Применимым решением оказалось сопоставление векторных представлений запроса и примеров интенга: подход на основе парных энкодеров обеспечивает высокую точность при десяти примерах на класс [9].

Короткие реплики и русский язык. Согласия, междометия и однословные реакции составляют значительную долю ходов в переписке и наиболее сложны для классификации: их значение определяется предшествующим контекстом. Для русского языка сложность возрастает вследствие свободного порядка слов и развитой морфологии, из-за чего модели, обученные на англоязычном материале, теряют точность при переносе. Адаптация многоязычных трансформеров даёт прирост, наиболее заметный на коротких фрагментах [10].

Обе задачи упираются в одно и то же слабое место, в короткую реплику. Эмоцию в ней не распознать без предыдущего хода, а намерение уж тем более. Например, фраза: «а можно?» без контекста не означает ничего. Поэтому системы, которые обрабатывают каждую реплику по отдельности, ведут себя довольно хорошо на длинных сообщениях и совершенно никак на переписке с короткими реакциями.

Заключение

В статье разобраны два способа описывать эмоцию, основные обучающие корпуса, три поколения архитектур и задача определения намерений вместе с отклонением чужих запросов.

В обеих задачах качество зависит от того, сколько разговора модель видит и знает ли она, кто и что сказал; увеличение параметров даёт не даёт сильного прироста в качестве. Отсюда практический вывод: проверять систему отдельно на коротких репликах и на чужих запросах, потому что там ошибки виднее всего. И пока те, кто размечает данные не могут прийти к какому-то одному выводу, предел точности зависит именно от качества данных, а не от самой модели.

Использованные источники:

1. Ekman P. An Argument for Basic Emotions // Cognition and Emotion. 1992. Vol. 6. № 3–4. P. 169–200.

2. Plutchik R. The Nature of Emotions // American Scientist. 2001. Vol. 89. № 4. P. 344–350.
3. Busso C., Bulut M., Lee C.-C. [et al.] IEMOCAP: Interactive Emotional Dyadic Motion Capture Database // Language Resources and Evaluation. 2008. Vol. 42. № 4. P. 335–359.
4. Poria S., Hazarika D., Majumder N. [et al.] MELD: A Multimodal Multi-Party Dataset for Emotion Recognition in Conversations // Proceedings of the 57th Annual Meeting of the ACL. 2019. P. 527–536.
5. Majumder N., Poria S., Hazarika D. [et al.] DialogueRNN: An Attentive RNN for Emotion Detection in Conversations // Proceedings of the AAAI Conference on Artificial Intelligence. 2019. Vol. 33. P. 6818–6825.
6. Ghosal D., Majumder N., Poria S. [et al.] DialogueGCN: A Graph Convolutional Neural Network for Emotion Recognition in Conversation // Proceedings of EMNLP-IJCNLP. 2019. P. 154–164.
7. Devlin J., Chang M.W., Lee K. [et al.] BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding // Proceedings of NAACL-HLT. 2019. P. 4171–4186.
8. Larson S., Mahendran A., Peper J.J. [et al.] An Evaluation Dataset for Intent Classification and Out-of-Scope Prediction // Proceedings of EMNLP-IJCNLP. 2019. P. 1311–1316.
9. Casanueva I., Temčinas T., Gerz D. [et al.] Efficient Intent Detection with Dual Sentence Encoders // Proceedings of the 2nd Workshop on NLP for Conversational AI. 2020. P. 38–45.
10. Kuratov Y., Arkhipov M. Adaptation of Deep Bidirectional Multilingual Transformers for Russian Language // Computational Linguistics and Intellectual Technologies. 2019. P. 333–340.