

Математическая модель оценки применимости интеллектуальных чат-ботов для изучения диалектов иностранных языков

Г.Э. Рего, Е.В. Рего

Петрозаводский государственный университет, Петрозаводск

Аннотация: В данной статье представлена математическая модель для оценки применимости интеллектуальных чат-ботов в изучении диалектов иностранных языков. Модель учитывает ключевые параметры чат-ботов и их способности адаптироваться к различным диалектам, оценивая качество ответов чат-бота как на основе определенных критериев, так и в целом. В рамках экспериментального исследования с немецкими диалектами оценивались такие чат-боты, как ChatGPT-3.5, GPT-4, YouChat, Bard, DeepSeek и Chatsonic. Результаты показали, что модели OpenAI (ChatGPT-3.5 и GPT-4) предоставляют наиболее широкие возможности для изучения диалектов, при этом ChatGPT-3.5 лучше всего справляется с баварским и австрийским диалектами, а YouChat - со швейцарским. Эти находки предоставляют практические рекомендации для выбора чат-ботов при изучении диалектов и являются основой для дальнейших исследований эффективности образовательных технологий на базе искусственного интеллекта.

Ключевые слова: большая языковая модель, чат-бот, оценка качества, изучение иностранного языка, технология искусственного интеллекта в образовании.

Введение

Сегодня, в эпоху глобализации и мгновенного обмена информацией, изучение иностранных языков становится неотъемлемой частью личного и профессионального развития. Однако, ключевым аспектом языкового многообразия, часто упускаемым из виду, является изучение отдельных диалектов. Это приобретает особое значение, поскольку диалекты обогащают языковой опыт и предоставляют уникальные культурные нюансы. Освоение диалектов в контексте изучения языков не только расширяет горизонты понимания языковой среды, но также предоставляет глубокое понимание и культурных особенностей местных сообществ, что становится особенно важным в международных и многоязычных общностях. В этом контексте, освоение диалектов приобретает актуальность не только для лингвистов, но и для широкого круга людей, стремящихся погрузиться в богатство языкового мира и локальных языковых традиций.

Для изучения иностранных языков могут использоваться технологии искусственного интеллекта [1, 2]. С увеличением числа доступных интеллектуальных чат-ботов вопрос эффективности их применения в конкретных образовательных задачах становится более сложным. На первый взгляд, наличие множества интеллектуальных чат-ботов предоставляет обширные возможности для выбора подходящего инструмента в обучении диалектам иностранных языков [3, 4]. В то же время возникает проблема оценки эффективности чат-ботов для учета индивидуальных особенностей диалектов иностранных языков.

Целью проведения данного исследования является построение математической модели оценки применимости интеллектуальных чат-ботов в контексте изучения диалектов иностранных языков, позволяющего повысить эффективность образовательного процесса.

Обзор научной литературы

Интеллектуальные чат-боты являются актуальной темой в контексте диалектов различных языков [5]. Появление ChatGPT (большая языковая модель, разработанная OpenAI) [6] привело к взрывному росту числа исследований на тему применения интеллектуальных чат-ботов в различных отраслях, в том числе, и в изучении иностранных языков [7]. В статье [7] исследователи приходят к выводу о необходимости тщательного изучения преимуществ, проблем и этических последствий использования интеллектуальных чат-ботов. Решающее значение имеют методы оценки, обеспечивающие безопасное использование, включая оценку точности и предвзятости ответов.

В исследовании [8] рассматривается идея смешанного метода обучения, которая заключается в интеграции чат-ботов в изучение языка. При этом снижается нагрузка преподавателя. Авторы приходят к выводу о

том, что успешная интеграция требует учета языковых и технологических особенностей, включая адаптивность и персонализацию.

Идея оценивать качество работы с интеллектуальными чат-ботами с точки зрения лингвистики не нова [9-11]. В исследовании [9] производительность пяти чат-ботов в различных грамматических категориях оценивалась с помощью балльной системы. Ответы оценивались на грамматическую правильность, то есть на связность, а также на сочетание того и другого. Использовалась бинарная система оценивания («один» за правильный ответ, «ноль» в случае любых ошибок). Такой подход облегчал сравнение для различных критериев оценивания, однако это приводило к тому, что грубые ошибки (например, нарушение порядка слов или орфографические ошибки) приравнивались к незначительным (неправильное использование заглавных букв и пунктуации).

В исследовании [10] содержится обзор литературы с 1990 года и статей с 2015 года, выделены атрибуты качества для чат-ботов и предлагаются подходы к обеспечению качества. Однако, многие критерии имеют качественный характер и их трудно описать в виде численного метода.

Также возможно применение статистических методов оценки качества интеллектуальных чат-ботов на основе данных, полученных в ходе опроса. В [11] применен десятишаговый метод для разработки шкалы измерений в области чат-ботов с применением технологий искусственного интеллекта. Недостатком такого метода можно считать необходимость применения больших усилий для получения оценки качества. В частности, в ходе исследования были опрошены 55 организаций в 17 странах мира.

Эти исследования показывают, что проблема оценки эффективности использования интеллектуальных чат-ботов в образовании, включая технологические аспекты, восприятие пользователей и эффективность обучения, является актуальной. Они предоставляют ценные практические и

теоретические выводы для дальнейших исследований и разработок в данной области. Однако, во всех этих исследованиях выполняется преимущественно качественная оценка работы интеллектуальных чат-ботов. В рамках данного исследования разработана математическая модель для количественной оценки результатов работы интеллектуальных чат-ботов.

Математическая модель оценки интеллектуальных чат-ботов

Обозначим число диалектов как K . Множество вопросов обозначим как $P = \{p_1, p_2, \dots, p_n\}$, где n – число вопросов на заданном диалекте. Введем множество критериев оценивания $C = \{c_1, c_2, \dots, c_m\}$, по которым оценивается работа чат-бота. Множество ответов обозначим как $A = \{a_1, a_2, \dots, a_n\}$. Также введем множество ошибок $E = \{e_{ij}\}$, где $i = 1, 2, \dots, m; j = 1, 2, \dots, n; k = 1, 2, \dots, K$; то есть каждая ошибка соответствует j -му ответу по i -му критерию.

Теперь опишем более подробно формирование элементов множества ошибок. Как уже было сказано выше, ошибки могут иметь разную стоимость, значит, этот аспект стоит учитывать в математической модели. Введем множество стоимостей для разных типов ошибок $V = \{v_1, v_2, \dots, v_l\}$. Каждая ошибка e_{ij} при ответе на j -й вопрос вычисляется, как сумма ошибок по критерию i . Если чат-бот допускает множество ошибок одного типа, то стоимость этих ошибок суммируется. Однако, величина e_{ij} не должна превышать единицу. Математически это можно выразить следующим образом:

$$e_{ij} = \begin{cases} \sum_{u=1}^l v_u * b_u, & \text{если } \sum_{u=1}^l v_u * b_u < 1 \\ 1 & \text{иначе} \end{cases} \quad (1),$$

где $0 < v_u \leq 1$ – стоимость ошибки u -о типа, b_u – число таких ошибок. Таким образом, областью определения e_{ij} является отрезок $[0; 1]$.

Общую оценку ответов на диалекте k по критерию c_i можно выразить, как функцию:

$$Q_k(c_i) = \frac{\sum_{j=1}^n (1 - e_{ij})}{n} \quad (2).$$

Областью определения величины $Q_k(c_i)$ является отрезок $[0; 1]$. Подобная нормировка может быть легко преобразована, например, в десятибалльную (путем умножения оценки на 10) или стобалльную (путем умножения оценки на 100) систему. В данном случае ответы на вопросы имеют равноценный характер. Однако, на практике это может быть и не так. В этом случае в формулу (2) необходимо ввести коэффициенты (веса) важности каждого вопроса. Подобные преобразования останутся за рамками данного исследования.

На основе сущностей, описанных выше, можно ввести общую оценку результативности чат-бота R , которая вычисляется по формуле:

$$R = \sum_{i=1}^m w_i * Q_k(c_i) \quad (3),$$

где $\sum_{i=1}^m w_i = 1$ и $w_1, w_2, \dots, w_m > 0$ – веса критериев.

Таким образом, областью определения величины R является отрезок $[0; 1]$. Представленная метод оценки в виде формулы (3) не является единственно верным. Например, альтернативным способом оценки может служить минимизация самой крупной ошибки. В этом случае качество работы чат-бота определяется как $1 - \max |e_{ij}|$.

Еще одним вопросом, который стоит осветить является выбор весов критериев в итоговой оценке w_i . В формуле (3) используется линейная функция, но может быть использована и любая другая (например, некоторые критерии могут пересекаться между собой и тогда имеет смысл выражать их влияние друг друга в виде перемножения (деления) весов и оценок).

Описанная математическая модель не заменяет экспертную оценку ответов чат-ботов. Именно эксперт определяет критерии оценивания, типы ошибок, их стоимости, а также задаваемые вопросы. В то же время, разработанная модель позволяет унифицировать процесс оценивания.

Апробация математической модели на примере диалектов немецкого языка

Для оценки применимости интеллектуальных чат-ботов для изучения диалектов иностранных языков необходимо выбрать критерии оценивания из множества C . В рамках данного исследования были использованы следующие критерии:

1. Орфография. В рамках данного критерия учитывается верное использование слов и выражений, которые могут быть употреблены только в определенном регионе или группе людей.
2. Правильное применение грамматики. В рамках данного критерия оценивается верное использование вариаций в грамматических конструкциях и правилах, таких, как изменения в форме глаголов, существительных или местоимений, а также особенности синтаксиса и структуры предложений.
3. Верная интерпретация слов, фраз и выражений в заданной теме, т.е. семантика. В рамках данного критерия оценивается верная интерпретация заданного вопроса, правильное понимание терминов, смысловая достоверность, точность логической структуры ответа.
4. Глубина ответа. В данном аспекте оценивалась детализация, конкретика и контекстуализация.

Таким образом в нашем случае множество $C = \{c_1 = \text{орфография}, c_2 = \text{грамматика}, c_3 = \text{семантика}, c_4 = \text{глубина}\}$.

В ходе исследования были оценены такие интеллектуальные чат-боты как ChatGPT-3.5, GPT-4, YouChat, Bard, DeepSeek и Chatsonic. Также, в экспериментальном исследовании предпринималась попытка использования отечественных чат-ботов YandexGPT 2 и GigaChat. В рамках проведенного исследования оценка данных чат-ботов на базе технологий искусственного интеллекта не была выполнена по причине отсутствия у них навыка

генерации текстов на немецком языке и его диалектах. Помимо этого, предпринималась попытка общения с чат-ботом DeepSeek. Данный чат-бот имеет навык общения на немецком языке, однако кодировку, в которой отсутствуют буквы немецкого алфавита «ä», «ö», «ü», «ß». В результате они заменяются на бессмысленный набор символов, затрудняющих чтение и понимание текста.

Несмотря на большое количество диалектов, общим для немецкоязычных стран (Германия, Австрия, Швейцария) является академический немецкий (Standardsprache) или литературный язык (Hochdeutsch). Также, в число используемых диалектов входят австро-баварские диалекты. Баварский (Bairisch), на котором говорят на территории Баварии [12], австрийский (Österreichisch) и швейцарский диалект немецкого языка (Schwiizerdütsch) [13]. Обладая разнообразием в лексике, фонетике и грамматике, германские диалекты могут вызывать трудности в восприятии даже для коренного населения Германии. Более того, эти диалекты классифицируются на несколько групп, которые, в свою очередь, подразделяются еще на несколько подгрупп [14].

Австрийский диалект немецкого языка значительно отличается от его литературного варианта. Например, род существительных, образование множественного числа и прочее. В Швейцарии используется диалект, имеющий особенности в грамматическом аспекте – специфика глагольного управления, употребления предлогов, а в лексике – специфическая окраска многих слов. Баварский диалект входит в группу верхненемецких диалектов. В первую очередь, баварский диалект отличается мягкостью произношения. Грамматика баварского также значительно отличается от академического немецкого. В баварском диалекте существуют только три падежа. Многие слова подвержены сильной ассимиляции, она может проявляться слиянием нескольких слов в одно [15].

Для каждого диалекта было составлено от 5 до 8 вопросов. Вопросы составлялись с применением грамматических правил и с использованием лексики данного диалекта. Тематика вопросов затрагивала особенности региона, где используется данный диалект. Например, был задан вопрос на баварском диалекте: «Wia vui Menschn lebn in Bayern?» (Сколько людей живет в Баварии?), с использованием лексики, которая характерна для данного диалекта (баварский диалект – wia; академический немецкий – wie). Также использовались грамматические особенности диалекта. (баварский диалект – lebn; академический немецкий – leben). С полным списком вопросов и ответов можно ознакомиться по ссылке (github.com/Messir1995/foreign-language-dialects).

Далее опишем множество ошибок. Все ошибки по типам делятся на незначительные, средней тяжести и грубые, стоимость которых описана во множестве $V = \{v_1=0.2, v_2=0.5, v_3=0.8\}$. Грамматические незначительные ошибки характеризуются отсутствием знаков пунктуации, прописных букв в начале предложений. Ошибками средней тяжести являются небольшие отклонения от грамматических норм используемого диалекта: неверное употребление падежей, артиклей, неверное образование множественного числа существительных, некорректное спряжение глаголов. Грубой ошибкой считается неиспользование грамматики заданного диалекта.

В орфографическом критерии незначительными ошибками является неверное написание слов, не затрудняющее понимание контекста. Ошибки средней тяжести заключались в неправильном написании слов, затрудняющем понимание контекста. Грубые ошибки характеризовались неиспользованием лексики, относящейся к данному диалекту, и её неверным употреблением.

В семантическом критерии незначительными считались ошибки в понимании деталей вопроса, которые не влияют на понимание общего

смысла. Ошибки средней тяжести заключались в неясности выраженной мысли, которая затрудняет понимание контекста. Грубыми считались ошибки, которые полностью меняют предполагаемый смысл, и фактические ошибки.

В критерии глубины незначительные ошибки – это отсутствие в ответе деталей, которые не влияют на раскрытие общего смысла. Ошибками средней тяжести являлись – недостаточная подробность в ответе и необходимость более точного раскрытия темы. Грубые ошибки в данном критерии заключались в краткости ответа, который не содержит информацию для общего понимания темы.

Результаты экспериментов представлены в таблице 1. Значения в строках «Орфография», «Грамматика», «Семантика», «Глубина» вычислялись по формуле (2). Значения в строках «Итоговая оценка» вычислялись по формуле (3). Значения в строке «Средняя итоговая оценка» вычислялись, как среднее арифметическое по значениям «Итоговая оценка».

Таблица № 1

Результаты тестирования работы чат-ботов с различными диалектами

Вариация немецкого языка	Критерий оценивания	ChatGPT-3.5	GPT 4	Bard (Gemini)	ChatSonic	YouChat
Hochdeutsch	Орфография	1	1	1	1	1
	Грамматика	1	1	1	1	1
	Семантика	1	1	1	0.97	1
	Глубина	0.91	0.91	1	0.71	0.97
	Итоговая оценка	0.98	0.98	1	0.92	0.99
Bairisch	Орфография	0.8	0.3	0.3	0.66	0.2
	Грамматика	0.8	0.3	0.275	0.66	0.2
	Семантика	1	1	0.875	1	1
	Глубина	0.8	0.88	0.8	0.74	0.6
	Итоговая оценка	0.85	0.62	0.56	0.77	0.5
Österreichisch	Орфография	0.77	0.2	0.2	0.31	0.2
	Грамматика	0.89	0.2	0.2	0.37	0.2
	Семантика	1	1	1	0.89	1
	Глубина	0.89	0.9	1	0.74	0.89
	Итоговая оценка	0.89	0.56	0.6	0.58	0.57
Schwiizerdütsch	Орфография	0.2	0.2	0.56	0.2	1
	Грамматика	0.2	0.2	0.56	0.2	1

	Семантика	1	0.96	1	1	1
	Глубина	0.84	0.84	0.96	0.72	0.7
	Итоговая оценка	0.56	0.55	0.77	0.42	0.93
Средняя итоговая оценка		0.82	0.82	0.73	0.67	0.75

На основе представленных в таблице данных, можно сделать следующие выводы. Все чаты, обладающие навыком общения на немецком языке, показали хорошие результаты при общении на академическом варианте немецкого языка. Лучшие результаты при общении на баварском и австрийском диалекте показал ChatGPT-3.5. Длина контекстуального окна данного чат-бота позволяла сохранять контекст на протяжении всего разговора. Почти всегда ChatGPT-3.5 верно применял лексику и грамматику используемого диалекта, правильно интерпретировал вопросы, давал достаточно глубокие ответы, которые полностью тему беседы.

Самый высокий показатель при общении на швейцарском диалекте немецкого языка был достигнут YouChat. Данный чат-бот имеет множество преимуществ при использовании швейцарского диалекта, таких, как безошибочное употребление грамматики и лексики и правильное понимание вопросов. Однако, YouChat дает очень краткие ответы, не предоставляя достаточного количества информации, для полного раскрытия заданной темы.

Стоит отметить, что Bard имеет несколько преимуществ над другими чатами. Он может использоваться в мультимодальном режиме, например, генерировать изображения в своих ответах, что помогает создавать более яркое впечатление о теме диалога. Помимо этого, чат Bard может озвучивать ответы. Данный навык у чат-бота значительно облегчает задачу изучения диалектов.

Заключение

В статье представлена математическая модель оценки применимости интеллектуальных чат-ботов в контексте изучения диалектов иностранных языков, основанная на анализе ключевых параметров и характеристик чат-

ботов, а также их способности адаптироваться к различным диалектам. В качестве параметров модели используются вопросы, ответы, критерии оценивания, типы и стоимости ошибок. Оценивается качество ответов чат-бота как по отдельным критериям, так и в целом. Для проверки эффективности предложенного метода проведено экспериментальное исследование на примере диалектов немецкого языка, что дает возможность оценить применимость предложенного подхода в реальных образовательных сценариях.

Анализ результатов применения разработанной математической модели для ведущих чат-ботов в изучении диалектов немецкого языка показал, что на данный наиболее широкий диапазон возможностей предоставляют модели компании OpenAI (ChatGPT-3.5 и GPT4). Полученные результаты позволяют предложить важные практические рекомендации для выбора интеллектуальных чат-ботов в области изучения диалектов иностранных языков, а также служить основой для дальнейших исследований в области оценки эффективности образовательных технологий на основе искусственного интеллекта.

Литература

1. Huang W., Hew K. F., Fryer L. K. Chatbots for language learning—Are they really useful? A systematic review of chatbot-supported language learning //Journal of Computer Assisted Learning. – 2022. – V. 38. – No. 1. – pp. 237-257.
2. Mageira K. et al. Educational AI chatbots for content and language integrated learning //Applied Sciences. – 2022. – V. 12. – No. 7. – pp. 3239-3255.
3. Hamad O., Hamdi A., Shaban K. Empathy and Persona of English vs. Arabic Chatbots: A Survey and Future Directions //International Conference on Text, Speech, and Dialogue. – Cham: Springer International Publishing, 2022. – pp. 525-537.

4. Alhassan N. A. et al. A novel framework for arabic dialect chatbot using machine learning //Computational Intelligence and Neuroscience. – 2022. – V. 2022. – pp. 1-11.

5. Chaves A. P. et al. Chatbots language design: The influence of language variation on user experience with tourist assistant chatbots //ACM Transactions on Computer-Human Interaction. – 2022. – V. 29. – No. 2. – pp. 1-38.

6. Lund B. D., Wang T. Chatting about ChatGPT: how may AI and GPT impact academia and libraries? //Library Hi Tech News. – 2023. – V. 40. – No. 3. – pp. 26-29.

7. Baskara R. et al. Exploring the implications of ChatGPT for language learning in higher education //Indonesian Journal of English Language Teaching and Applied Linguistics. – 2023. – V. 7. – No. 2. – pp. 343-358.

8. Belda-Medina J., Calvo-Ferrer J. R. Using chatbots as AI conversational partners in language learning //Applied Sciences. – 2022. – V. 12. – No. 17. – pp. 8427.

9. Coniam D. The linguistic accuracy of chatbots: usability from an ESL perspective //Text & Talk. – 2014. – V. 34. – No. 5. – pp. 545-567.

10. Radziwill N. M., Benton M. C. Evaluating quality of chatbots and intelligent conversational agents //arXiv preprint arXiv:1704.04579. – 2017.

11. Chen Q. et al. Classifying and measuring the service quality of AI chatbot in frontline service //Journal of Business Research. – 2022. – V. 145. – pp. 552-568.

12. Шевченко Е. В. К проблеме распространения и развития баварского диалекта немецкого языка //Вестник Московского государственного лингвистического университета. – 2012. – №. 646. – С. 211-214.

13. Тиньков Т. В. Основные варианты и диалекты немецкого языка //Гуманитарное образование: история, традиции, перспективы. Том Выпуск

5. Елец: Елецкий государственный университет им. И.А. Бунина. – 2017. – С. 63-67. elibrary.ru/item.asp?id=30550543&selid=30615278

14. Васильева В. А. Сравнительный анализ диалектов в немецком языке //Концептуальные пути развития гуманитарных и социальных наук: сборник материалов XV-ой международной очно-заочной научно-практической конференции, 25 января, 2023. – Москва: Издательство. – 2023. – С. 44.

15. Салихова Д. Б., Кухтенков А. П. Особенности баварского диалекта //Современные проблемы развития техники, экономики и общества. Материалы II Международной научно-практической очно-заочной конференции. Издательство: "Рóкета Союз". – 2017. – С. 200-202. elibrary.ru/item.asp?id=29172281

References

1. Huang W., Hew K. F., Fryer L. K. Journal of Computer Assisted Learning. 2022. No. 1. pp. 237-257.

2. Mageira K. et al. Applied Sciences. 2022. №. 7. pp. 3239-3255.

3. Hamad O., Hamdi A., Shaban K. International Conference on Text, Speech, and Dialogue. Cham: Springer International Publishing. 2022. pp. 525-537.

4. Alhassan N. A. et al. Computational Intelligence and Neuroscience. 2022. pp. 1-11.

5. Chaves A. P. et al. ACM Transactions on Computer-Human Interaction. 2022. No 2. pp. 1-38.

6. Lund B. D., Wang T. Library Hi Tech News. 2023. No 3. pp. 26-29.

7. Baskara R. et al. Indonesian Journal of English Language Teaching and Applied Linguistics. 2023. No. 2. pp. 343-358.

8. Belda-Medina J., Calvo-Ferrer J. R. Using chatbots as AI conversational partners in language learning. Applied Sciences. 2022. No 17. p. 8427.

9. Coniam D. The linguistic accuracy of chatbots: usability from an ESL perspective. Text & Talk. 2014. No 5. pp. 545-567.
10. Radziwill N. M., Benton M. C. Evaluating quality of chatbots and intelligent conversational agents. arXiv preprint arXiv:1704.04579. 2017.
11. Chen Q. et al. Classifying and measuring the service quality of AI chatbot in frontline service. Journal of Business Research. 2022. pp. 552-568.
12. Shevchenko E. V. Vestnik Moskovskogo gosudarstvennogo lingvisticheskogo universiteta. 2012. № 646. pp. 211-214.
13. Tin`kov T. V. Gumanitarnoe obrazovanie: istoriya, tradicii, perspektivy`. Tom Vy`pusk 5. Elec: Eleczkij gosudarstvenny`j universitet im. I.A. Bunina. 2017. pp. 63-67.
14. Vasil`eva V. A. Sravnitel`ny`j analiz dialektov v nemeczkom yazy`ke [Comparative analysis of dialects in German]. Konceptual`ny`e puti razvitiya gumanitarny`x i social`ny`x nauk: sbornik materialov XV-oj mezhdunarodnoj ochno-zaochnoj nauchno-prakticheskoy konferencii, 25 yanvarya, Moskva: Izdatel`stvo. 2023. P. 44.
15. Salixova D. B., Kuxtenkov A. P. Osobennosti bavarskogo dialekta [Features of the Bavarian dialect]. Sovremenny`e problemy` razvitiya texniki, e`konomiki i obshhestva. Materialy` II Mezhdunarodnoj nauchno-prakticheskoy ochno-zaochnoj konferencii. Izdatel`stvo: "Róketa Soyuz". 2017. Pp. 200-202.

Дата поступления: 25.11.2023

Дата публикации: 11.01.2024