

О ПОДХОДАХ К УЧЕТУ ТЕМПЕРАМЕНТА В ДИАЛОГОВЫХ ПРОГРАММАХ

Волкова Л.Л., Комаров Н.С.
МГТУ им. Н. Э. Баумана
+7 (499) 263-68-08, liliya@bmstu.ru, kom.niki@mail.ru

Статья рассматривает понятие темперамента человека в свете задачи переноса особенностей реакций человека с заданным темпераментом на диалоговые программы. Описаны отличительные черты и типологии темпераментов человека. Представлены существующие подходы к моделированию темперамента: вероятностный, Карпова, на основе логического вывода, на основе языковой модели. Сделан вывод о перспективности аналитических детерминированных моделей, в т.ч. ввиду их интерпретируемости. Указано, что интерес представляет верификация модели на основе сетей Петри и опросов респондентов по итогам взаимодействия с программой, учитывающей модельный темперамент при формировании реакций на входные стимулы в беседе.

Ключевые слова: темперамент, диалоговые системы, аффективная составляющая коммуникации, человеко-машинные интерфейсы.

On approaches to considering temperament in dialogue software. Volkova L.L., Komarov N.S. Bauman Moscow State Technical University.

This article considers the notion of human temperament in the focus of the problem of transferring human reactions peculiarities into dialogue programs with condition of a given temperament. Typologies and distinctive traits of human temperaments are presented along with existing approaches to modelling temperaments: probabilistic, Karpov and those based on either logical inference or language models. Analytical determined models are pointed out as prospective, namely due to their interpretability. Petri net-based verification of a model is of interest, as well as that grounded on respondents surveys upon interaction with a program which accounts for a model temperament when forming reactions to input stimuli in conversation.

Keywords: temperament, dialogue systems, affective communication, human-machina interfaces.

Введение

Темперамент — характеристика индивида со стороны его динамических особенностей: интенсивности, скорости, темпа, ритма психических процессов и состояний. Основными компонентами данной характеристики являются активность и эмоциональность. Активность поведения характеризует степень энергичности, стремительности, быстроты и, наоборот, медлительности, инертности, а эмоциональность — особенности протекания эмоций, чувств, настроений и их качество [1]. Понятие темперамента было введено древнегреческим врачом Гиппократом [2], его теория называется гуморальной. Согласно этой теории, особенности поведения человека зависят от состояния его организма, в связи с чем Гиппократ выделял четыре основных типа темперамента: холерик (вспыльчивый, энергичный, эмоциональный), флегматик (спокойный, серьезный, малоэмоциональный), меланхолик (замкнутый, ранимый, задумчивый), сангвиник (жизнерадостный, активный).

Теория Гиппократа получила дальнейшее развитие в работах И. П. Павлова и Г. Айзенка, учитывающий психофизиологический аспект [3]. И. П. Павлов [4] описал физиологическую основу темперамента как тип высшей нервной деятельности, определяемый соотношением основных свойств нервной системы: силы, уравновешенности и подвижности процессов возбуждения и торможения. Слабый тип нервной системы наблюдается у меланхолика, сильный — у остальных темпераментов. Из них холерик обладает неуравновешенной нервной системой, сангвиник и флегматик — уравновешенной. При этом нервная система флегматика инертна, сангвиника — подвижна. Г. Айзенк в своей теории [5] выделял два основных, независимых фактора темперамента (две оси): интроверсия-экстраверсия и эмоциональная стабильность-нестабильность (нейротизм). На основе комбинаций значений этих факторов можно описать четыре типа темпераментов, как показано в таблице 1.

Таблица 1 – Типология темпераментов по Г. Айзенку

Темперамент	Интроверсия-экстраверсия	Эмоциональная стабильность
Меланхолик	Интроверсия	Низкая
Холерик	Экстраверсия	Низкая
Сангвиник	Экстраверсия	Высокая
Флегматик	Интроверсия	Высокая

Диалоговая система — это автоматизированная человеко-машинная система, взаимодействующая с пользователем на ограниченном естественном языке [6]. В базе своей данный класс программ ориентирован на понимание текстов и зачастую на диалоговое взаимодействие с оператором [7]. Такие

системы могут быть реализованы в виде голосовых ассистентов, роботов, чат-ботов, виртуальных помощников и иных приложений [8]. Отдельный класс диалоговых систем — роботы- и программы-собеседники, моделирующие аффективную составляющую восприятия входных воздействий и формирования ответных реакций на смысл входного высказывания [9–11] по аналогии с тем, как это делают люди. Такие проекты показывают определенные подвижки в достоверности поведения роботов и человеко-машинных интерфейсов: алгоритмически сформированное и воспроизведенное поведение оценивается респондентами как приятное и человекоподобное [12], по данным агрегации обратной связи о восприятии человеко-машинного взаимодействия в режиме повседневного и целеориентированного диалога. Замечено, что людям свойственно относиться к роботам как к участникам социальной активности и подсознательно присваивать им персональные черты [13]. Под персональными понимают те характеристики человека, которые находят отражение в устойчивых паттернах чувств, мышления и поведения [14]. Введение функциональности для воспроизведения свойственных различным темпераментам черт и особенностей реакций — одно из перспективных направлений в развитии диалоговых систем [15].

В данной работе будут рассмотрены известные алгоритмические подходы к моделированию темперамента в диалоговых системах, явно описывающие поведение системы в ответ на входные воздействия: подход на основе логического вывода, вероятностный подход, подход Карпова, подход на основе языковой модели, подход на основе большой языковой модели.

Подход на основе логического вывода

Данный подход использовался для разработки таких диалоговых систем, как ELIZA [16] и SHRDLU [17], и может быть адаптирован к применению в поставленной задаче. Кроме этого, современные голосовые ассистенты бывают реализованы гибридным методом: некоторые модули их программ реализованы с использованием логического вывода. Пример: AliMe Assist [18] проверяет запрос пользователя на соответствие шаблонам, и только если соответствие найти не удаётся, передаёт запрос классификатору, использующему алгоритмы машинного обучения. Данный подход основан на принципах логического программирования как, например, в языке Prolog [19]. Ответ на реплику пользователя ищется путём подбора применимого к запросу правила. Особенности данного подхода: (1) все параметры модели принимают значения из конечного множества, так как эти значения должны быть использованы в программе; (2) при одних и тех же параметрах модели и входном воздействии пользователя программа будет выдавать одинаковые результаты, т.е. подход детерминированный.

Вероятностный подход

Развивая идею предыдущего подхода, можно разработать алгоритм, который на внешнее воздействие формирует несколько вариантов возможного поведения — коммуникативных целей (КЦ), как в проекте Ф-2 [9]. Выбор КЦ из списка может быть осуществлён вероятностно, причём «вес» каждой КЦ зависит от моделируемого темперамента, возможно, комбинированного. Суть подхода сводится к вычислению значений функции вероятности, зависящей от внешнего воздействия на систему, от реакции системы и от внутренних параметров системы, на всей области определения функции. Данный подход не накладывает ограничений на модель темперамента. В работах [20–21] используется данный подход.

Подход Карпова

В. Э. Карпов в [22] использовал модель темперамента, основанную на типах нервной системы, выделенных Павловым. В данном подходе предполагается, что поведенческие реакции описываются двумя факторами: силой и балансом нервных процессов возбуждения и торможения. Описанная В. Э. Карповым модель темперамента включает параметр возбуждения, отвечающий за задержку реакции, и параметр торможения, отвечающий за длительность нахождения системы в заданном состоянии. Эти параметры используются как обратная связь в системе управления робота, по аналогии с примерами смены состояний решающего умственную задачу человека: человек проходит череду сменяющих друг друга состояний — от отвлечённого к повышенному состоянию внимания [23]. Типология темпераментов в терминах возбуждения и торможения представлена в таблице 2 и совпадает с типологией Г. Айзенка (см. табл. 1). В частности, исследована применимость различных темпераментов для агентов с различными ролями в многоагентной системе [23].

Таблица 2 – Типология темпераментов в терминах возбуждения и торможения

Темперамент	Возбуждение	Торможение
Меланхолик	Низкий	Низкий
Холерик	Высокий	Низкий
Сангвиник	Высокий	Высокий
Флегматик	Низкий	Высокий

Подход на основе языковой модели

Машинное обучение (англ. Machine Learning, ML) — это подраздел искусственного интеллекта, который изучает методы построения моделей и алгоритмов, которые позволяют алгоритмам «учиться», то есть извлекать знания из данных без явного программирования этих знаний [24]. Существует три основных типа машинного обучения [25]:

- обучение с учителем использует набор данных, организованных в виде коллекции размеченных пар вход–выход, где выход представляет целевой результат;
- обучение без учителя использует набор данных без разметки;
- обучение с подкреплением использует данные о состоянии «среды», а также «поощрение» и «наказание» за принятые решения (добавление и снятие очков целевой функции в аддитивном случае, домножение на коэффициент в мультипликативном случае).

Языковая модель — это статистическая модель, описывающая распределение вероятностей по последовательностям слов (или иных языковых единиц или их групп, например, n-грамм) в языке [26]. В данном подходе используется языковая модель на основе нейронной сети. Если нейронная сеть будет «обучена» на подходящем наборе данных, она сможет формировать ответы на запросы пользователя, генерируя текст с необходимой стилистической окраской. Стиль этого текста может быть сформирован на основе как имитации авторского стиля конкретного человека или литературного персонажа, так и особенностей, придуманных именно для заданной диалоговой системы. Таким образом, темперамент возможно задать, отобрав для обучающей выборки тексты авторов с ярко выраженным темпераментом.

Стоит отметить, что в данном подходе не обязательно «обучать» нейронную сеть с нуля. Вместо этого можно воспользоваться уже существующей языковой моделью и «дообучить» её, используя набор данных, специализированный под конкретную задачу. Данный подход используется в голосовых ассистентах, таких как Яндекс Алиса [27] и СяоАйс [28].

Разработчики Яндекс Алисы позиционируют её следующим образом: «Алиса — голосовой помощник с характером. Она умна, любопытна и любит пошутить. Алисе важно помогать людям, но она знает себе цену и не потерпит оскорблений. При этом и не опустится до хамства, когда ей грубят» [29]. СяоАйс — это китайский чат-бот, разработанный компанией Microsoft. Он способен поддерживать разговор с людьми, отвечать на вопросы, развлекать, шутить и проявлять эмпатию. СяоАйс может динамически распознавать эмоции и вовлекать пользователя в долгие разговоры, предлагая соответствующие межличностные реакции [30].

В составе данного подхода возможно отдельно выделить работу с большой языковой моделью (англ. Large Language Model, LLM). Это языковая модель, состоящая из нейронной сети с большой размерностью вектора параметров (миллиарды весовых коэффициентов и более), обученной на большом количестве неразмеченного текста с использованием обучения без учителя [31]. Преимуществом является отсутствие необходимости «дообучать» модель на дополнительном наборе данных. Вместо этого необходимо в запрос пользователя добавить дополнительную информацию, описывающую стиль ответа, соответствующий моделируемому темпераменту. Однако, не все существующие LLM могут быть использованы в данном подходе, так как не все они обладают необходимой глубиной обучения. На данный момент, реализовать данный подход позволяет GPT-4 [32]. В качестве примера может служить чат-бот Bing Chat от Microsoft [33], основанный на GPT-4. В нём реализован «режим знаменитости», в котором чат-бот подражает знаменитостям, копируя их манеру речи, «коронные» фразы, темперамент.

Однако у данных методов множество недостатков, от «галлюцинаций» до несоответствия состава и численности подсистем имеющимся у человека (по данным картирующих мозг нейробиологов), от возможности обходить запреты на конкретную тематику через запрос «представь, что...» до доверчивости к данным, которые должны бы априорно отбираться экспертами на предмет достоверности, отсутствия односторонности, пристрастности, агрессивности, нарушения законодательства [34]. Часть проблем связана с необходимостью предварительной оценки тональности текстов обучающей выборки, в том числе с учётом иронии [35–36], оценки тональности текста и выделения фрагментов текста, направленных на унижение человеческого достоинства и разжигание розни [37]. Также есть значительное число проблем с неоднозначностями в естественном языке, от наличия или отсутствия решения которых критически зависит качество результата обработки текстов [38].

Сравнение существующих подходов к моделированию темпераментов

Для сравнения рассмотренных подходов (см. табл. 3) выбраны следующие критерии:

- 1) детерминированность подхода — подход детерминированный, если для идентичных начальных условий и входных воздействий будет получен один и тот же результат;
- 2) базовый механизм, положенный в основу подхода;
- 3) использование машинного обучения.

Таблица 3 – Сводная таблица подходов к моделированию темпераментов

Подход \ Критерий	1	2	3
На основе логического вывода	да	правила	нет
Вероятностный	нет	вероятности выбора реакций при заданном темпераменте	нет
Подход Карпова	да	регулирование параметров возбуждения и торможения	нет
На основе языковой модели	нет	вероятности определяются статистикой на основе корпусных данных	да, возможно дообучение

Заключение

Ввиду множественности недостатков подхода на основе языковых моделей, в частности, не-правильности данного подхода (в смысле отсутствия по построению строгих правил и ограничений), данный подход категорически не рекомендуется к применению для поставленной задачи на уровне выше демонстрационных студенческих лабораторных работ. Наибольший интерес представляют аналитические подходы, позволяющие верифицировать результат и обеспечивать интерпретируемость [39] поведения программы на алгоритмическом уровне, сиречь первые три рассмотренных подхода по отдельности или в совокупности, в гибридном формате.

Отдельный интерес представляет процедура верификация сценариев реагирования диалоговой системы. Она может быть основана на опросе респондентов [15], может зиждиться на сетях Петри [40].

Литература

1. Карпенко Л.А., Ярошенко М.Г., Петровский А.В. (ред.). Краткий психологический словарь. — Ростов-на-Дону: «ФЕНИКС», 1998. — 512 с.
2. История психологии: От античности до наших дней: Учеб. для вузов по спец. «Психология» / А. Н. Ждан. — М.: Изд-во МГУ, 1990. — 366 с.
3. Грушева О.А. Теории темперамента / Научные исследования и разработки молодых ученых. — 2014. — № 1. — С. 39–42.
4. Физиология нервной системы: Избранные труды : В 4 вып. / И. М. Сеченов [и др.]; Под общ. ред. акад. К. М. Быкова. — Москва: Медгиз, 1952. — С. 606.
5. Психология индивидуальных различий: учебное пособие для студентов высших учебных заведений / Е. П. Ильин. — Санкт-Петербург и др.: Питер, 2011. — 700 с.
6. Мильчин А.Э., Издательский словарь-справочник. — М.: ОЛМА-Пресс, 2003. — 558 с.
7. Волкова Л.Л. О диалоговых системах и классификации их функциональности // Проблемы лингвистики и лингводидактики в неязыковом вузе: 5-я Международная научно-практическая конференция (Москва, 15–16 декабря 2022 года): сборник материалов конференции : в 2 т. / Федеральное государственное бюджетное образовательное учреждение высшего образования «Московский государственный технический университет имени Н. Э. Баумана (национальный исследовательский университет)»; Факультет «Лингвистика» МГТУ им. Н. Э. Баумана; Ассоциация технических университетов России и Китая. — М.: Издательство МГТУ им. Н. Э. Баумана, 2023. — Т. 1. — С. 43–49.
8. Виноград Т. Программа, понимающая естественный язык: Пер. с англ. / Терри Виноград; Пер. С. С. Калитин, Ред. пер. Г. Е. Поздняк. — М.: Мир, 1976. — 298 с.
9. Kotov A., Arinkin N., Filatov A., Kivva K., Zaidelman L., Zinina A. Cognitive architecture for a companion robot: speech comprehension and real-world awareness // Velichkovsky B.M., Balaban P.M., Ushakov V.L. (eds) Advances in Cognitive Research, Artificial Intelligence and Neuroinformatics. Proc. 9th Int. Conf. on Cognitive Sciences, Intercogsci-2020, October 10–16, 2020, Moscow, Russia. AISC 1358. — Cham: Springer, 2021. — Pp. 641–647.
10. Malkina M., Zinina A., Arinkin N., Kotov A. Multimodal hedges for companion robots: a politeness strategy or an emotional expression? // Selegey, V.P., et al. (eds.) Computational Linguistics and Intellectual Technologies: Papers from the Annual International Conference “Dialogue”, issue 22. — Moscow: RSUH, 2023. — Pp. 319–326.
11. Polianskaya L., Volkova L. A CycleGAN-based method for translating recordings of interjections // Kravets A.G., Shcherbakov M.V., Groumpos P.P. (eds) Creativity in Intelligent Technologies and Data Science. CIT&DS 2023. CCIS 1909. — Cham: Springer, 2023. — Pp. 83–97.
12. Velichkovsky B.M., Kotov A., Arinkin N., Zaidelman L., Zinina A., Kivva K. From social gaze to indirect speech constructions: how to induce the impression that your companion robot is a conscious creature // Applied Sciences. — 2021. — Vol. 11(21):10255. — Pp. 1–19.
13. De Graaf M.M.A., Ben Allouch S. Expectation setting and personality attribution in HRI // HRI, ACM/IEEE Int. Conf. on Human-Robot Interaction. — Piscataway: IEEE, 2014, pp. 144–145.
14. Pervin, L. A., John, O. P. Personality theory and research. — New York: Wiley, 1997. — 578 p.

15. Volkova L., Kotov A., Ignatev A. Crowdsourcing-based approbation of communicative behaviour elements on the F-2 robot: perception peculiarities according to respondents / *Studies in Computational Intelligence*, LNCS 1130. — Cham: Springer, 2024. — Pp. 932–945.
16. Weizenbaum J. ELIZA — a computer program for the study of natural language communication between man and machine / *Communications of the ACM*. — 1966. — Vol. 9. — №. 1. — Pp. 36–45.
17. Winograd T., Flores F. *Understanding Computers and Cognition: A New Foundation for Design*. — Norwood: Ablex Pub. Comp., 1986. — 224 p.
18. Li F.-L., Qi M., Chen H., Wang X., Gao X., Huang J., Ren J., Zhao Z., Zhao W., Wang L., Jin G., Chu W. AliMe assist: An intelligent assistant for creating an innovative e-commerce experience // *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*. — 2017. — New York: ACM, 2017. — Pp. 2495–2498.
19. Information technology, Programming languages, Prolog [эл. ресурс]. URL: <https://www.iso.org/standard/21413.html> (дата обращения 12.03.2024).
20. Калачин С. В. Нечеткое моделирование механизма передачи панического состояния среди людей с различными видами темперамента / *Компьютерные исследования и моделирование*. — 2021. — Т. 13. — №. 5. — С. 1079–1092.
21. Савельев А.П., Чугунов М.Н., Глотов С.В., Еналеева С.А., Никифорова И.А., Шкрабак Р.В. Совершенствование методики расчета пожарного риска на объектах с массовым пребыванием людей / *Аграрный научный журнал*. — 2022. — №. 12. — С. 95–99.
22. Карпов В.Э. Методы группового управления искусственными агентами на основе биологически инспирированных моделей поведения: автореф. дис. на соискание степени доктора технических наук. — Москва: 2022. — 48 с.
23. Карпов В.Э., Карпова И.П., Кулинич А.А. Социальные сообщества роботов. — М.: ЛЕНАНД, 2017. — 352 с.
24. Domingos P. A few useful things to know about machine learning / *Communications of the ACM*. — 2012. — Vol. 55. — №. 10. — Pp. 78–87.
25. Бурков А. *Машинное обучение без лишних слов*. — СПб.: Питер, 2020. — 192 с.
26. Jurafsky D., Martin J.H. N-gram Language Models // *Speech and Language Processing*. Draft of December 30, 2020. [эл. ресурс]. URL: https://web.stanford.edu/~jurafsky/slp3/old_dec20/3.pdf (дата обращения 31.05.2024).
27. Яндекс Алиса [эл. ресурс] URL: <https://alice.yandex.ru> (дата обращения 14.03.2024).
28. XiaoIce [эл. ресурс] URL: <https://news.microsoft.com/ru-ru/kakbudto-po-telefonu-xiaoice-sotsialny-j-chat-bot-microsoft-zapushhenny-j-v-kitae-obshhaetsya-pochti-po-chelovecheski/> (дата обращения 14.03.2024).
29. Кухновец К. С. Принцип работы голосового помощника Яндекс Алиса [эл. ресурс] URL: <https://rep.bntu.by/handle/data/141446> (дата обращения 14.03.2024).
30. Shum H.-Y., He X.-D., Li D. From Eliza to XiaoIce: challenges and opportunities with social chatbots / *Frontiers of Information Technology & Electronic Engineering*. — 2018. — № 19. — Pp. 10–26.
31. Что такое большие языковые модели (LLM) [эл. ресурс]. URL: <https://aws.amazon.com/ru/what-is/large-language-model> (дата обращения 14.03.2024).
32. OpenAI's GPT-4 [эл. ресурс]. URL: <https://openai.com/gpt-4> (дата обращения 14.03.2024).
33. Your Everyday AI Companion | Microsoft Bing [эл. ресурс]. URL: <https://www.microsoft.com/ru-ru/bing?form=MA13FV> (дата обращения 14.03.2024).
34. Dietterich, T.G.: What's wrong with LLMs and what we should be building instead (keynote talk at ValgrAI Scientific Council Forum 2023, July 4th 2023) [эл. ресурс]. URL: <https://www.youtube.com/watch?v=cEyHsMzbZBs> (дата обращения 2024/07/15).
35. Волков Е. А. Метод определения тональности отзывов на русском языке с учётом иронии. Выпускная квалификационная работа бакалавра. МГТУ им. Н.Э. Баумана, 2021 г.
36. Dolbin A., Rozaliev V., Orlova Yu., Ulyev A. Automating the Detection of Sarcastic Statements in Natural Language Text / *Advances in Automation* — 2020. — ? — Pp.1185–1194.
37. Хеламбаге Г. Классификатор токсичных комментариев с использованием обработки естественного языка и машинного обучения. Выпускная квалификационная работа бакалавра. МГТУ им. Н.Э. Баумана, 2021 г.
38. Большакова Е.И., Клышинский Э.С. Ландэ Д.В., Носков А.А., Пескова О.В., Ягунова Е.В. Автоматическая обработка текстов на естественном языке и компьютерная лингвистика. Учебное пособие. — М: МИЭМ, 2011. — 272 с.
39. Рудаков И.В. Методика иерархического исследования сложных дискретных структур / *Наука и образование*. — 2012. — № 06. — С. 251–260. — DOI: 10.7463/0612.0370230.
40. Рудаков И.В., Медведев В.О. Алгоритм верификации программного обеспечения с помощью иерархических сетей Петри / *Наука и бизнес: пути развития*. — 2019. — № 1(91). — С. 74–77.