

УДК 510.58

# Вероятностные методы для сравнительного анализа и оптимизации сценариев и алгоритмов управления речевым диалогом

М. П. Фархадов\*, Н. В. Петухова\*

*\* Лаборатория автоматизированных систем массового обслуживания и обработки сигналов,*

*Институт проблем управления им. В.А. Трапезникова РАН,  
ул. Профсоюзная, д. 65, Москва, Россия, 117997*

**Аннотация.** В статье вероятностными методами исследуются основные характеристики человеко-машинного диалога в речевых диалоговых системах: вероятность успешного завершения диалога и оценка продолжительности диалогов с различными процедурами выявления и корректировки ошибок. Выполнен анализ различных вариантов диалогов и получены рекомендации для оптимизации сценариев диалогов с учетом надежности распознавателя.

**Ключевые слова:** вероятностные методы, сценарии и алгоритмы управления речевым диалогом, вероятность успешного завершения диалога, продолжительность диалога, речевой интерфейс.

## 1. Введение

Автоматизированные диалоговые системы на основе речевых технологий все активнее внедряются в повседневную жизнь людей. Чаще всего их используют для получения информации, при бронировании или для передачи голосовых команд. Центральная роль в речевых диалоговых системах принадлежит модулю управления диалогом. Его функцией является принятие решения относительно того, по какой из ветвей будет развиваться диалог после завершения каждого человеко-машинного взаимодействия. Алгоритмы поведения проектируются разработчиком и основываются на оптимизации определенной функции с учетом текущих значений ряда параметров. Основную проблему при проектировании управления диалогом составляет необходимость учета ошибок распознавания. Несмотря на то, что надежность распознавателей значительно выросла за последние годы, ошибки все равно будут возникать. Они являются следствием и внешнего шума, и индивидуальности произношения клиентов, и наличия трудно различаемых по соотношению слов в любом языке. Единственный надежный способ обнаружить и исправить ошибки — это обращение к процедуре переспроса, когда система после распознавания сказанного клиентом произносит свою гипотезу и просит клиента подтвердить или отвергнуть ее. Однако переспрос имеет и негативную сторону: он связан с потерей времени и может вызывать раздражение клиента. Со всем *не безразлично*, в каком месте диалога использовать переспрос и

как его производить. В статье вероятностными методами исследуются варианты сценариев диалога с применением различных процедур обнаружения и коррекции ошибок распознавания.

## 2. Сравнение сценариев диалога

В речевых диалоговых системах взаимодействие с пользователем осуществляется путем последовательных циклов вопрос-ответ, в которых каждый вопрос системы построен так, чтобы получить в ответе клиента некоторую информацию. Ответ клиента обрабатывается распознавателем. В зависимости от результатов обработки меняется внутреннее состояние диалога и формируется новый вопрос клиенту или следует обращение к базе данных или внешней системе.

Вероятностные методы достаточно широко используются для анализа и оптимизации автоматизированных диалоговых систем, построенных на основе речевых технологий. В работе [1] поведение системы представляется марковским процессом принятия решений. В работе [2] авторы рассматривали проблему управления диалогом с учетом возможных ошибок распознавания. Выбор предлагаемого им алгоритма продолжения диалога основан на использовании частично наблюдаемого марковского процесса принятия решений. В работе [3] поведение системы представляется в виде графа, и модуль управления выбирает следующее действие системы, основываясь на текущем состоянии и в соответствии с прописанной стратегией, оптимизирующей целевую функцию. Подобные подходы доказали свою эффективность в простых приложениях. Однако они становятся слишком сложными для конструирования правил в сложных сценариях.

В последние годы были предложены новые подходы, которые используют статистические методы и машинное обучение для определения оптимальных действий системы в ходе диалога. Обучение с подкреплением является в настоящее время наиболее популярным методом, используемым для оптимизации диалога на основе накопленных данных. В работах [4, 5] приведен обзор публикаций и результаты некоторых применений обучения с подкреплением в диалоговых системах. Показано, что такие системы демонстрируют неплохие результаты.

Однако полноценное применение машинного обучения на этапе разработки систем и на ранних стадиях эксплуатации невозможно из-за отсутствия данных для обучения. Были предложены различные модели пользователей для тренировки моделей, однако поведение реальных пользователей оказывается, как правило, отличным от моделируемого [6]. Другой подход состоит в создании обобщенной базы данных [7]. Здесь есть свои проблемы, например, предполагается равенство энтропий всех атрибутов.

Целью данной работы является анализ основных сценариев диалога и получение ряда оценок и рекомендаций, которые помогли бы разработчикам речевых диалоговых систем спроектировать оптимальный

сценарий диалога в зависимости от надежностных параметров используемых речевых блоков.

Для решения этой задачи выполнена *классификация диалогов*, в основу которой положены следующие признаки: типы элементов диалога - простые элементы и составные элементы; местоположение процедур выявления и корректировки ошибок - в *процессе* формирования запроса, *после окончания* формирования запроса; способ реализации процедур выявления и корректировки ошибок - *покаждому элементу, единым блоком, путем опроса*.

Методика выполнения сравнительного анализа вариантов состоит в использовании модифицированного применительно к данной задаче принципа *квантификации*, когда варианты сравниваются на множестве *одинаковых типовых количественных значений* ряда параметров. Например, сравнительная оценка вероятности успешного завершения диалога осуществляется при *одинаковых значениях вероятностей* правильного распознавания речевых блоков и *одинаковом допустимом* числе переспросов для сравниваемых вариантов сценариев [8, 9, 10].

Необходимость в таких допущениях диктуется большой размерностью моделей и широким диапазоном изменения параметров. *Квантификация является достаточно популярным методом количественного анализа качества интерфейсов*.

В качестве характеристик для сравнения вариантов выбраны два главных количественных критерия оценки качества речевого диалогового человеко-машинного взаимодействия: *вероятность успешного завершения диалога и продолжительность диалога*.

Поиск оптимального сценария осуществляется на основе минимизации оценок продолжительности диалога с обеспечением заданной вероятности его успешного завершения.

*Оценка вероятности успешного завершения диалога.*

Для оценки *вероятности успешного завершения диалога* вначале определяются вероятности правильного распознавания *элементов диалога*, а затем выводятся формулы вероятности успешного завершения *диалогов*, отличающихся типами элементов и процедурами выявления и корректировки ошибок.

Под *простым элементом* диалога понимается взаимодействие между системой и клиентом в виде «вопрос-ответ», в ходе которого выявляется один атрибут запроса.

Под *составным элементом* диалога понимается взаимодействие, при котором определяются несколько атрибутов запроса с использованием разных речевых блоков. Использование составных элементов в диалоге вызвано обычно стремлением сократить продолжительность диалога.

Предельным вариантом составного элемента является «*монологовое*» взаимодействие, при котором в ответе клиента на вопрос системы присутствуют все запрашиваемые атрибуты.

Вероятность правильного распознавания элемента диалога зависит от надежности соответствующего данному шагу диалога речевого блока (блоков) и от влияния человеческого фактора. Надежности речевых

блоков оцениваются экспериментально путем их тестирования. Влияние человеческого фактора в процессе переспросов может быть уменьшено за счет выдачи примеров и подсказок. Далее показаны формулы для оценки вероятности правильного распознавания простого и составного элементов.

Вероятность правильного распознавания простого элемента диалога при  $n$  переспросах  $P_n$  определяется формулой:

$$P_n = \rho_0 + \sum_{j=1}^n \rho_{0j} \prod_{k=0}^{j-1} (1 - \rho_{0k}), n = 1, 2, 3$$

где  $\rho_0$  - номинальная вероятность правильного распознавания элемента диалога,  $\rho_{0k}$  - вероятность правильного распознавания при  $k$ -том переспросе. В условиях однородности процесса распознавания, когда вероятность распознавания при каждом переспросе равна  $\rho_0$ , вероятность правильного распознавания элемента диалога представляется усеченным геометрическим распределением:

$$P_n = \sum_{j=0}^n \rho_0 (1 - \rho_0)^j = \rho_0 \cdot \frac{1 - (1 - \rho_0)^{n+1}}{1 - (1 - \rho_0)} = 1 - (1 - \rho_0)^{n+1}, n = 0, 1, 2, 3$$

Величина, вычисленная по данной формуле, представляет собой нижнюю оценку вероятности правильного распознавания элемента диалога при  $n$  переспросах.

Из формулы видно, что при  $\rho_0 > 0,7$  один-два переспроса могут обеспечить вероятность распознавания элемента диалога более 0,99.

Составной элемент содержит  $M$  смысловых составляющих, распознаваемых самостоятельно с вероятностью  $P_{0i}$ ,  $i = 1, M$ . Вероятность  $P_{0c}$  правильного распознавания составного элемента диалога равна

$$P_{0c} = \prod_{i=0}^M P_{0i}$$

Для составного элемента, как можно получить из формулы, также достаточно 1 или 2 переспросов, а наилучшие результаты по вероятности распознавания имеет вариант выявления и корректировки ошибок по каждой составляющей элемента.

Вероятность успешного завершения диалога  $P_{\partial n}$ , состоящего из комбинации  $(N - N_c)$  простых и  $N_c$  составных элементов, каждый из которых может переспрашиваться  $n$  раз, определяется соотношением:

$$P_{\partial n} = \prod_{i=1}^{N-N_c} P_{n_i} \cdot \prod_{j=1}^{N_c} P_{c_{n_j}}$$

### Продолжительность диалога

Для анализа диалогов по продолжительности введены следующие разновидности оценок:

- предельные оценки: оценка снизу, оценка сверху;
- средняя оценка.

Нижняя оценка вычисляется в предположении отсутствия неправильно распознанных элементов и характеризует алгоритм управления диалогом при идеальной системе распознавания. Оценка сверху характеризует максимально возможную продолжительность диалога при уменьшении вероятности распознавания элемента диалога до установленного предела.

Формулы для расчета продолжительности диалога учитывают в том числе и временные затраты на процедуры выявления и корректировки ошибок. Были рассмотрены следующие варианты этих процедур: по каждому шагу диалога, после завершения диалога путем переспроса всех атрибутов, после завершения диалога и прочтения распознанного запроса путем выявления ошибочных атрибутов и их корректировки.

Важно отметить, что выявление и корректировка ошибок распознавания во время диалога приводит к автоматизации накопления информации об ошибках в лог файлах. В результате создается база данных, которая может быть использована как при исправлении ошибок в словарях, так и при применении различных методов машинного обучения для разработки эффективных стратегий управления диалогом на последующих стадиях развития проекта.

Ниже приведены общие формулы для вычисления оценок продолжительности диалога из  $N$  элементов: нижней оценки  $T_H$ , оценки сверху  $T_B$  и среднего времени  $T_{cp}$ .

$$T_H = \left( \sum_{i=1}^N L_i \right) \cdot \tau$$

$$T_B = \left[ \left( \sum_{i=1}^N L_i \right) \cdot K_{n_{\partial on}} + \sum_{i=1}^N Q_i + \left( \sum_{i=1}^N G_i \right) \cdot K_{n_{\partial on}} \right] \cdot \tau$$

$$T_{cp} = \left[ \sum_{m=0}^N (1 - p_0)^m \cdot p_0^{N-m} \cdot C_N^m \cdot \left[ L_{cp} \cdot (N - m) + \right. \right. \\ \left. \left. + (L_{cp} \cdot K_{n_{\partial on}} + Q_{cp} + G_{cp} \cdot K_{n_{\partial on}}) \cdot m \right] \right] \cdot \tau$$

где  $\tau$  - временной квант, затрачиваемый на произнесение одного среднестатистического слова,

$L_i$  - количество временных квантов, затрачиваемых на  $i$ -тый элемент диалога при правильном распознавании,

$Q_i$  - количество временных квантов, затрачиваемых на  $i$ -тый элемент диалога при переспросе,

$G_i$  - количество временных квантов, дополнительно затрачиваемых на  $i$ -тый элемент диалога при переспросе непонятых реплик клиента,

$K_{переспрос}$  - коэффициент, показывающий, во сколько раз увеличивается время, затрачиваемое на один переспрашиваемый элемент диалога при допустимом числе переспросов.

Сравнение вариантов организации диалогов произведено для безразмерных величин, которые получены делением обеих частей соотношений для временных оценок на величину  $\tau$ . Эти безразмерные величины названы соответственно коэффициентами нижней, верхней и средней оценок продолжительности диалога.

### 3. Результаты анализа и рекомендации

В соответствии с классификацией методов речевого взаимодействия были проанализированы следующие варианты организации диалогов:

- диалог из простых элементов;
- диалог из простых и составных элементов;
- режим «монолога».

Для каждого варианта были рассмотрены различные процедуры выявления и корректировки ошибок. Варианты сравнивались по критериям продолжительности диалога при обеспечении заданной вероятности успешного распознавания.

Выполненный анализ позволил сформулировать следующие рекомендации для разработчиков речевых диалоговых систем:

- В диалогах из простых элементов лучше производить переспросы по каждому элементу.

- Использование составных элементов улучшает временные характеристики для стратегий введения переспросов после формирования запроса. Улучшение тем больше, чем выше надежность элементов.

- Режим монолога имеет лучшие характеристики, чем другие, но только при достаточно надежных речевых блоках.

- При отсутствии конкретных данных о надежности распознавания рекомендуется строить диалог из простых элементов с проверкой правильности распознавания и корректировкой ошибок по каждому элементу диалога.

Полученные результаты также позволяют:

- определять необходимое число переспросов;
- определять среднее время, затрачиваемое на элемент диалога;
- вычислять вероятности успешного завершения диалогов различной структуры;
- анализировать и сравнивать между собой различные стратегии управления диалогом;

- производить обоснованный выбор сценариев и алгоритмов управления диалогом для конкретных задач на основе разработанной методики.

#### 4. Заключение

Проектирование сценариев для речевых диалоговых систем во многом остается искусством. Человек привык к речевому взаимодействию с другим человеком и поэтому непроизвольно ожидает от речевых систем такого же «разговора». Однако распознавание речи и ее понимание пока еще далеко не свободны от ошибок. Разработчик речевых приложений должен считаться с этими факторами. Поэтому для него важны обоснованные рекомендации, которыми он может воспользоваться на ранних этапах своей работы, чтобы создать успешно функционирующую систему, пусть и с упрощенным диалогом типа вопрос-ответ и с простыми процедурами выявления и корректировки ошибок распознавания. Данное исследование было нацелено на получение таких рекомендаций путем применения вероятностных моделей для анализа таких важных характеристик диалоговых систем как вероятность успешного завершения диалога и оценка его продолжительности. Применение различных вариантов процедур выявления и корректировки ошибок сопровождается автоматическим накоплением сведений об ошибках распознавания в лог файлах, что может быть использовано для улучшения качества работы систем на основе применения стратегий машинного обучения для управления диалогом.

#### Литература

1. *Young S.J.* Probabilistic methods in spoken-dialogue systems // Philosophical Transactions: Mathematical, Physical and Engineering Sciences. — 2000. — Vol. 358, no. 1769. - P. 1389–1402.
2. *Roy N., Pineau J., Thrun S.* Spoken Dialogue Management Using Probabilistic Reasoning, Proceedings of the 38th Annual Meeting of the Association for Computational Linguistics (ACL 2000). Hong Kong, Oct. 2000. — P. 93–100.
3. *Pieraccini R., Huerta J.* Where do we go from here? research and commercial spoken dialog systems // 6th SIGdial Workshop on Discourse and Dialogue, 2005. — P. 1-10.
4. *Young S., Gaci M., Thomson B., Williams J.D.* POMDP-based statistical spoken dialog systems: A review // Proceedings of the IEEE. — 2013.— Vol. PP, no. 99. — P. 1-20.
5. *Lemon O., Pietquin O.* Machine Learning for Spoken Dialogue Systems // Proceedings of the 10th European Conference on Speech Communication and Technologies (Interspeech'07), 2007. — P. 2685-2688.

6. *Crook P.A., Keizer S., Wang Z., Tang W., Lemon O.* Real user evaluation of a pomdp spoken dialogue system using automatic belief compression // *Computer Speech & Language*. — 2014. — Vol. 28, no. 4. — P. 873–887.
7. *Polifroni J., Walker M.A.* Intensional summaries as cooperative responses in dialogue: Automation and evaluation // *ACL*. Citeseer, 2008. — P. 479–487.
8. *Биллик Р.В., Жожикашвили В.А., Петухова Н.В., Фархадов М.П.* Анализ речевого интерфейса в интерактивных сервисных системах. Часть I // *Автоматика и телемеханика*. — 2009. — No. 2. — С. 80–89.
9. *Биллик Р.В., Жожикашвили В.А., Петухова Н.В., Фархадов М.П.* Анализ речевого интерфейса в интерактивных сервисных системах. Часть II // *Автоматика и телемеханика*. — 2009. — No. 3. — С. 97–113.
10. *Биллик Р.В., Мясоедова Э.П., Петухова Н.В., Фархадов М.П.* Анализ речевого интерфейса при взаимодействии клиента с автоматизированной системой массового обслуживания. — М.: МАКС Пресс, 2007. — 112 с.

UDC 510.58

## Probabilistic methods for comparative analysis and optimization of scenarios and algorithms for managing speech dialogue

M. P. Farkhadov\*, N. V. Petukhova\*

*\* Laboratory of automated queuing systems and signal processing,  
V.A. Trapeznikov Institute of Control Sciences RAS,  
Profsoyuznaya str. 65, Moscow, Russia, 117997*

In the article, the basic characteristics of human-machine dialogue in speech dialogue systems are investigated by probabilistic methods: the probability of successful completion of a dialogue and evaluation of the duration of dialogues with various procedures for identifying and correcting errors. The analysis of various variants of dialogues is made and recommendations are received for optimizing the scenarios of the dialogues taking into account the reliability of the recognizer.

**Keywords:** probabilistic methods, scenarios and algorithms for managing speech dialogue, probability of successful completion of the dialogue, duration of the dialogue, speech interface.