

УДК 004.934

НЕКОТОРЫЕ АСПЕКТЫ ИССЛЕДОВАНИЯ СИСТЕМ РАСПОЗНАВАНИЯ РЕЧИ В ТЕЛЕФОННЫХ СЛУЖБАХ ПОДДЕРЖКИ

П.В. Балакшин, Г.Ю. Петров

Представлены проблемы телефонных служб поддержки клиентов (call-центров) крупных компаний. Определены преимущества и недостатки применения систем распознавания речи в работе таких центров. Показана важность метрики FCR (First Call Resolution) и особенности ее вычисления в системах автоматического распознавания речи.

Ключевые слова: распознавание речи, телефонная служба поддержки, call-центр, FCR.

Введение

В настоящее время активно ведутся исследования в области речевого взаимодействия «человек – компьютер», идет поиск новых параметров исходных данных, новых методов, возможных ограничений и объединений, позволяющих повысить точность и другие показатели распознавания речи компьютерными системами. Актуальность данных исследований подтверждается большим количеством международных научно-технических конференций, посвященных этим проблемам (Interspeech (1993–2011 г.г.), SpeCom (1996–2011 г.г.)), крупными программно-техническими разработками последних лет (Dragon Naturally Speaking, встроенное речевое управление в операционной системе Vista, речевой поиск Google).

Автоматическое распознавание речи находит все более широкое применение в работе call-центров (call center) – телефонных служб поддержки клиентов крупных компаний, деятельность которых сводится к оперативному решению телефонных запросов от клиентов или потенциальных потребителей. В небольших компаниях с этим могут справиться несколько человек, но для крупных организаций (например, страховых или транспортных компаний, финансовых учреждений) такое положение дел неприемлемо (табл. 1).

В настоящее время ряд таких центров рассматривают возможность частичной замены операторов на компьютерные устройства, способные распознавать речь, обрабатывать, анализировать ее и выдавать осмысленный и необходимый ответ клиенту [2].

| Вероятность услышать сигнал «Занято», % | Количество линий |
|---|------------------|
| 90 | 3 |
| 80 | 6 |
| 70 | 8 |
| | |
| 10 | 28 |
| 5 | 31 |
| 4 | 32 |
| 3 | 33 |
| 2 | 34 |
| 1 | 36 |

Таблица 1. Необходимое количество линий при 500 входящих вызовах в час [1]

Целью настоящей работы являлось исследование и оптимизация систем автоматического распознавания речи в работе call-центров.

Построение модели системы распознавания

Следует подчеркнуть, что общение оператора телефонной службы поддержки и клиента происходит в виде живой беседы. Следовательно, распознавание речи должно происходить если не в реальном, то в квазиреальном масштабе времени. Преимуществами использования систем распознавания речи в телефонных службах поддержки являются [3–6]:

- существенное снижение времени ожидания (handle time), что позволит снизить затраты на работу;
- сокращение времени звонка в 1,5–2 раза за счет сокращения времени ввода информации оператором;
- возможность автоматического ответа на простые вопросы, освобождая время работы операторов для работы со сложными звонками;
- возможность работы с клиентами в круглосуточном режиме (включая праздники);
- быстрая верификация голоса клиента во время ответа на один–два несложных вопроса, что особенно актуально в банковском секторе для защиты от краж персональных карт и секретных документов;
- возможность работы с большим числом коротких звонков (телефонная служба поддержки букмекерских компаний);
- возможность замены сложной и допускающей частые ошибки IVR-системы (Interactive Voice Response), работающей в тональном режиме;

- возможность использовать распознавание речи в качестве дополнительного источника информации не только при разговоре, но и при дальнейшем анализе работы call-центра. В частности, данный анализ помогает повысить ключевую метрику FCR – разрешение проблемы (вопроса) за один звонок. Это приводит к снижению числа повторных звонков и повышению удовлетворенности клиентов одновременно, что в свою очередь приводит к снижению операционных расходов.

К основным недостаткам использования систем распознавания речи следует отнести:

- невозможность распознавания с точностью 100% из-за большого количества различных особенностей произношения, таких как диалект, дефекты речи, эмоции, и ассоциативного восприятия речи человеком;
- очень большая начальная стоимость внедрения системы распознавания речи.

Предлагаемая авторами модель системы распознавания речи включает в себя блок обратной связи на изменение размера речевого словаря после обработки полученной информации (рис. 1). С одной стороны, меньший словарь обеспечит более высокую скорость распознавания. С другой стороны, существует вероятность распознать редкий вопрос с очень плохой точностью. По этой причине данный параметр, а именно, размер речевого словаря системы распознавания, очень важен: от него зависит скорость вовлечения распознавания речи в работу телефонной службы поддержки.

В рамках проведенных исследований по предлагаемой модели [7] было определено, что для русского языка словарь размером 2500 слов является достаточным для покрытия более 95% информации усредненной предметной области. Исходя из этого, не менее 95% информации будет распознаваться с точностью, присущей используемой модели.

В итоге, возможность модификации размера речевого словаря внутри самой системы в режиме реального (квазиреального) времени становится преимуществом использования распознавания речи. Особенно очевидно данное преимущество в работе телефонных служб поддержки, так как каждая из них имеет строго конкретизированную предметную область.

Данные результаты для русского языка получены впервые. Однако они находятся в соответствии с аналогичными исследованиями по английскому языку. Исследователь Ч. Огден, сформировавший в 1930 г. BASIC English, показал, что при строгой грамматике необходимо 850 слов. В дальнейшем он увеличил словарь до 1350 слов, а «стандартный» английский язык, использующийся в упрощенной версии Википедии, состоит из 2000 слов.

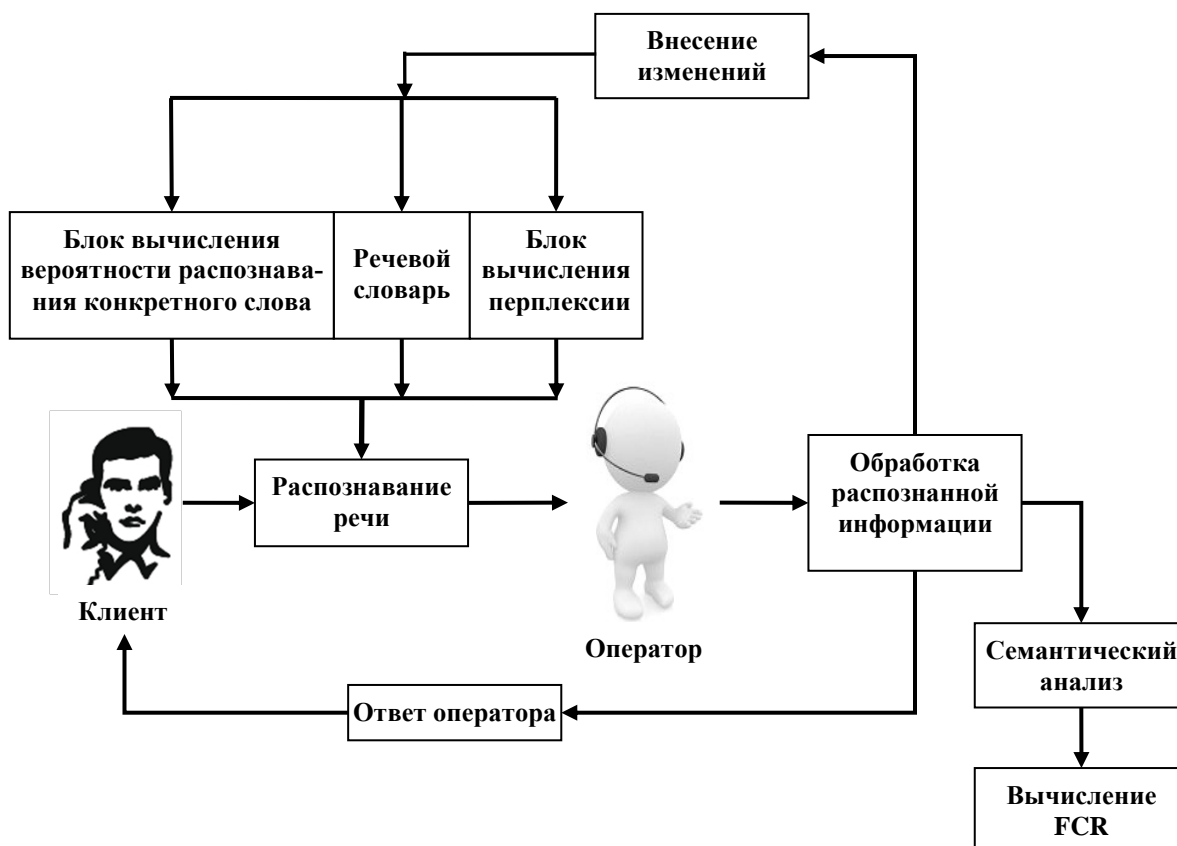


Рис. 1. Модель системы распознавания речи

Дополнительной особенностью модели является блок обратной связи на изменение вероятности распознавания конкретного слова после обработки полученной информации (рис. 1). Таким образом, блок вычисления вероятности распознавания конкретного слова работает в паре с речевым словарем. Благодаря этому появляется возможность увеличивать вероятность распознавания как заранее определенных, так и новых слов.

Выделим еще одну важную особенность предложенной модели. Авторами предложено учитывать в модели связность слов, или так называемую перплексию. Связность учитывает количество слов, которые логически могут встретиться после данного слова. Именно связность повышает точность распознавания речи, поскольку вопросы, задаваемые пользователями, имеют достаточно четкий синтаксис, что особенно ярко выражено в английском языке. Для этого выделен специальный блок вычисления перплексии (рис. 1), используемый во время распознавания, в который с помощью блока обратной связи вносятся изменения по результатам обработки распознаваемой информации.

Перплексия вычислялась по формуле [8]

$$PPL = 2^{\sum_{i=1}^N \frac{1}{N} \log_2 q(x_i)}$$

где N – количество состояний модели q ; x_i – состояние модели. Чем меньше значение перплексии, тем лучше модель предсказывает следующее слово.

Для проведения исследований за основу были взята языковая модель, использованная при реализации системы Decipher [9]. В данной системе при размере словаря в 998 слов в модели, использующей пары слов, перплексия равна 60, а в модели, использующей отдельные слова, перплексия равна 998, т.е. размеру всего словаря. В качестве исходных данных для исследования была составлена языковая модель на основе речевого словаря телефонной службы поддержки системного интегратора «Системные решения» (г. Москва). Словарь, составленный по результатам прослушивания звонков, содержал 1867 слов. Результаты исследования аналогичны системе Decipher (табл. 2).

Таким образом, лучшие результаты были получены при использовании групп из трех слов. Дополнительным подтверждением корректности результатов могут служить исследования о величине перплексии в зависимости от предметной области (табл. 3).

| Количество связанных слов в языковой модели | 1 | 2 | 3 | 4 | 5 | 6 |
|---|------|-----|----|-----|-----|-----|
| Перплексия | 1867 | 243 | 96 | 180 | 491 | 702 |

Таблица 2. Соотношение связанных слов и перплексии в языковой модели

| Предметная область | Перплексия |
|-------------------------------|------------|
| Радиология | 20 |
| Неотложная медицинская помощь | 60 |
| Журналистика | 105 |
| Общий английский | 247 |

Таблица 3. Перплексия триграммной языковой модели для различных предметных областей [10]

Как было отмечено ранее, преимуществом использования распознавания речи в телефонных службах поддержки является как само распознавание, так и дальнейший анализ полученной информации. В соответствии с результатами последних исследований call-центров [6, 8, 11], проведенных для различных отраслей бизнеса, 33% от общего числа звонков являются повторными (рис. 2).

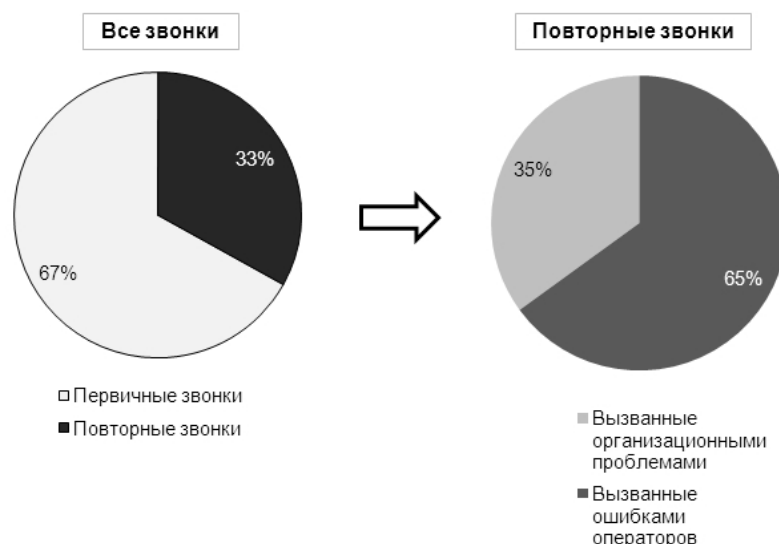


Рис. 2. Исследование причин возникновения повторных звонков

Особенности вычисления и использования метрики FCR

Наши исследования показали, что предлагаемая модель может с успехом использоваться для снижения количества повторных звонков. Это приводит к снижению общего количества звонков и снижению операционных расходов телефонной службы поддержки. Указанные изменения неразрывно связаны с улучшением основной метрики FCR, определяющей работу всей телефонной службы поддержки. Данная метрика оценивает процент разрешения проблемы (вопроса) за один звонок и вычисляется следующим образом [12]:

$$FCR_k = \frac{E_k}{N_k}, \quad FCR_K = \frac{\sum_{k=1}^K E_k}{\sum_{k=1}^K N_k},$$

где K – общее количество причин звонка; k – элемент из множества причин K ; N_k – количество цепочек звонков по k -ой причине; E_k – количество цепочек звонков по k -ой причине, состоящих из одного звонка. Тогда общее значение FCR можно определить как среднее из всех FCR_k .

Более высокое значение метрики является показателем более успешной работы. По итогам опроса руководителей 54 телефонных служб поддержки Великобритании 81% из них считает, что FCR – это именно тот показатель, который может объективно оценить работу как всей телефонной службы поддержки, так и каждого оператора (или группу операторов), в частности [3].

Существует несколько методов определения метрики FCR, среди которых выделяют [13]:

- текущий контроль за соединениями;
- опрос с помощью интерактивной IVR-системы по окончании звонка;
- телефонный опрос спустя один-три дня после первоначального звонка (survey – анкетирование, оценка, экспертиза).

В предлагаемом методе по итогам распознавания причины звонка и клиента данная информация передается в блок вычисления FCR (рис. 1), где на основании подсчета повторных звонков от каждого клиента по каждой причине вычисляется значение метрики FCR для каждой причины (FCR_k), а затем и итоговая FCR. Текущее значение FCR составляет 65–75%, что влечет довольно большие расходы на работу технической службы поддержки.

Рассмотрим более подробно, почему именно распознавание речи помогает повысить метрику FCR. Одной из главных проблем call-центров, которую возможно решить с применением автоматического распознавания речи, является определение причины звонка. Оператор либо выбирает эту информацию из списка уже известных причин, либо кратко записывает в текстовое поле информацию, услышанную от клиента. Но для части звонков определить причину звонка все равно невозможно, во многом из-за некорректной информации, введенной в специальное поле вручную. Так, в call-центре одного американского оператора связи процент звонков без причины составляет 10–12, а в call-центре одного довольно крупного американского банка – 27–30.

Показано, что в системе, в которой осуществлена предлагаемая авторами модель, количество звонков с конкретной причиной увеличивается, а за счет дальнейшего анализа повышается и значение метрики FCR до 72–79%. Исследованные звонки были распознаны со скоростью около 95 слов в минуту. Данный показатель является неплохим результатом, так как средняя скорость русской речи составляет 105–125 слов. Однако часто клиенты телефонной службы поддержки произносят свою речь с большим количеством эмоций, заметно ускоряя скорость речи и приближая ее к показателям английской речи – 140–150 слов в минуту.

Отметим еще один положительный момент. Введение в модель системы распознавания речи блока семантического анализа (рис. 1) позволяет повысить именно семантическую точность распознавания и улучшить значение метрики FCR.

Рассмотрим, например, следующие звонки:

- «Банкомат съел мою карту»;
- «Банкомат не вернул мне карту»;
- «Банкомат не принимает карту»;
- «Карта не выдается банкоматом обратно».

В данном случае важно то, что банкомат имеет проблемы при работе с картой. Таким образом, как причина звонка будет регистрироваться именно эта информация, а не ее семантически менее важные детали. За счет этого для определения метрики FCR будет использоваться уникальная причина звонка вместо различных схожих, т.е. семантический анализ выполняет некоторую предобработку данных.

Благодаря использованию семантического анализа в данных экспериментах метрика FCR увеличивается до 78–86%.

Заключение

В работе показано, что применение компьютерных систем распознавания речи, основанных на модели с модификацией словаря, и дальнейший анализ полученной информации позволяют повысить производительность телефонных служб поддержки и снизить затраты на их работу. Предложенная разработка помогает повысить ключевую метрику FCR. Это приводит к снижению числа повторных звонков, значительно снижает операционные расходы, повышает удовлетворенность клиентов.

Работа выполнена при поддержке гранта Правительства Петербурга № 3.11/04-06/50.

Литература

1. Официальный сайт компании Град Телеком [Электронный ресурс]. – Режим доступа: <http://www.gradtelecom.ru/ru/papers/callcenter/>, св. Яз. рус. (дата обращения 07.04.2010).
2. Хижинский Д. Распознавание речи взорвет рынок call-центров // cnews.com: сетевой журн. 22.05.2007. [Электронный ресурс]. – Режим доступа: <http://www.cnews.ru/reviews/index.shtml?2007/05/22/251191>, св. Яз. рус. (дата обращения: 11.05.2010).
3. Contact centres rate call efficiency over customer service // callcentrehelper.com: сетевой журн. 07.04.2010. [Электронный ресурс]. – Режим доступа: <http://www.callcentrehelper.com/contact-centres-rate-call-efficiency-over-customer-service-9921.htm#more-9921>, св. Яз. англ. (дата обращения: 28.04.2010).
4. How is speech recognition applied in call centers? [Электронный ресурс]. – Режим доступа: <http://www.glocall.com/voice-recognition/speech-recognition-in-call-centers.html>, св. Яз. англ. (дата обращения 19.12.2010).
5. Feldon B. The top five uses of speech recognition technology [Электронный ресурс]. – Режим доступа: <http://www.callcentrehelper.com/the-top-five-uses-of-speech-recognition-technology-1536.htm>, св. Яз. англ. (дата обращения 19.12.2010).
6. Neustein A. Advanced in Speech Recognition. Mobile Environments, Call Centers and Clinics. – Springer Science. – 2010. – 370 p.
7. Балакшин П.В., Тропченко А.Ю. Выбор размера словаря при реализации тестовой системы распознавания речи // Интеллектуальные и информационные системы. Материалы Всероссийской научно-технической конференции. – Тула: Изд-во ТулГУ, 2009. – С. 51–54.
8. Call Centre Monitoring // callcentrehelper.com: сетевой журн. 2010 [Электронный ресурс]. – Режим доступа: <http://www.callcentrehelper.com/call-centre-monitoring-16.htm>, св. Яз. англ. (дата обращения: 22.04.2010).
9. Cohen M., Murveit H., Bernstein J., Price P., Weintraub M. The Decipher speech recognition system // IEEE ICASSP. – Albuquerque, 1990. – P. 77–80.
10. Roukos S., Battista Varile Giovanni, Antonio Zampolli. Language Representation. Survey of the State of the Art in Human Language Technology. – 1995 [Электронный ресурс]. – Режим доступа: <http://cslu.cse.ogi.edu/HLTsurvey/HLTsurvey.html>, св. Яз. англ. (дата обращения: 05.09.2011).
11. Increase First Call Resolution: Eliminate the Top 10 Agent Errors Causing Repeat Calls // crmxchange.com: сетевая библиография: 2008 [Электронный ресурс]. – Режим доступа: <http://www.crmxchange.com/offer/enkatadec08.asp>, св. Яз. англ. (дата обращения: 28.12.2009).

12. Fluss D. FCR Improves Contact Center Performance // g-cem.org: сетевой журн. 2010 [Электронный ресурс]. – Режим доступа: http://www.gccrm.com/eng/content_details.jsp?contentid=2245&subjectid=107, св. Яз. англ. (дата обращения: 13.04.2011).
13. Стандарт EN 15838:2009 [Электронный ресурс]. – Режим доступа: http://www.iccci.ru/iccci/call_centers/en_15838/, св. Яз. рус. (дата обращения: 13.04.2011).

Балакиин Павел Валерьевич – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, ассистент, pvbalakshin@gmail.com

Петров Григорий Юрьевич – Энката Технолоджис, Инк., кандидат экономических наук, директор отдела внедрения, gpetrov@enkata.com