

Определение необходимого размера словаря для систем автоматического распознавания речи телефонных служб поддержки клиентов

П.В. Балакишин

*Санкт-Петербургский национальный исследовательский университет
информационных технологий, механики и оптики*

Аннотация: Кратко описана необходимость систем распознавания речи в телефонных служб поддержки клиентов (call-центров) крупных компаний. Описано текущее положение для русскоязычных систем. Представлена система распознавания речи с возможностью модификации речевого словаря. Определены факторы, влияющие на размер словаря. Подтверждено значение необходимого размера словаря русского языка конкретной предметной области.

Ключевые слова: распознавание речи, телефонная служба поддержки клиентов, call-центр, речевой словарь, словоформа, перплексия.

Введение

Исследования в области распознавания речи стали производиться вместе с распространением первых компьютеров, так как первоначальный текстовый интерфейс работы оператора с ЭВМ не обладал ни разумной скоростью, ни удобством. Более того, учёные продолжают искать новые параметры исходных данных, новые методы, всевозможные ограничения и объединения, позволяющие повысить точность и другие показатели распознавания речи. Это обусловлено тем фактом, что голосовой интерфейс на языке пользователя – наиболее натуральное, удобное, эффективное и экономичное решение.

Развитие столь актуальной для общества области подтверждается рядом международных научно-технических конференций (Диалог, SpeCom, Interspeech), разработкой коммерческих программно-аппаратных средств (например, Dragon Naturally Speaking, речевого поиск Google), публикациями в международных журналах.

Около 10 лет назад автоматическое распознавание речи нашло применение в достаточно новой для России области: в работе колл-центров (англ. *call center*) – телефонных служб поддержки клиентов. Некоторые из

таких служб ведут свои разработки для частичной замены операторов на определённые программно-аппаратные средства, которые смогут распознавать речь, обработать её, проанализировать и обеспечить звонящего клиента осмысленным и нужным ответом [1,2]. Целью данной работы является определение нижней границы размера словаря в системах распознавания речи, реализованных в контакт-центрах.

Предлагаемая система распознавания

К настоящему моменту уже реализованы коммерческие системы с неплохими результатами. Но большинство из них имеют английский интерфейс и, следовательно, используют речевой словарь английского языка. Также стоит отметить единичные, но успешные решения для испанского, французского, китайского и арабских языков (например, система распознавания речи контакт-центра аэропорта г. Дубай). Прежде всего такая ситуация обусловлена экономическими и политическими аспектами развития речевых технологий, так как средства, вложенные в развитие технологий автоматической обработки английской речи, быстро окупились. Дополнительные трудности при разработке систем автоматического распознавания русской речи возникают вследствие того, что русский язык относится к группе флективных языков: в нём присутствует большая вариативность, отсутствие строгих грамматических конструкций при построении предложений, многочисленные правила словообразования, фонетического представления слов и расстановки ударений с большим количеством исключений. Всё это влияет и на размер необходимого словаря распознавания.

Важной особенностью при реализации любого словаря является его предметная область [3]. С одной стороны, чем меньше словарь, тем более высокую скорость распознавания он может обеспечивать. С другой стороны, существует вероятность распознавания редкого слова или вопроса с очень

плохой точностью. Поэтому размер речевого словаря системы распознавания является одним из ключевых, так как от него зависит скорость внедрения распознавания речи в работу телефонной службы поддержки.

Предлагаемая система распознавания речи [4] включает в себя блок обратной связи на изменение размера речевого словаря на основании дополнительной обработки полученной от клиента информации (рис. 1).

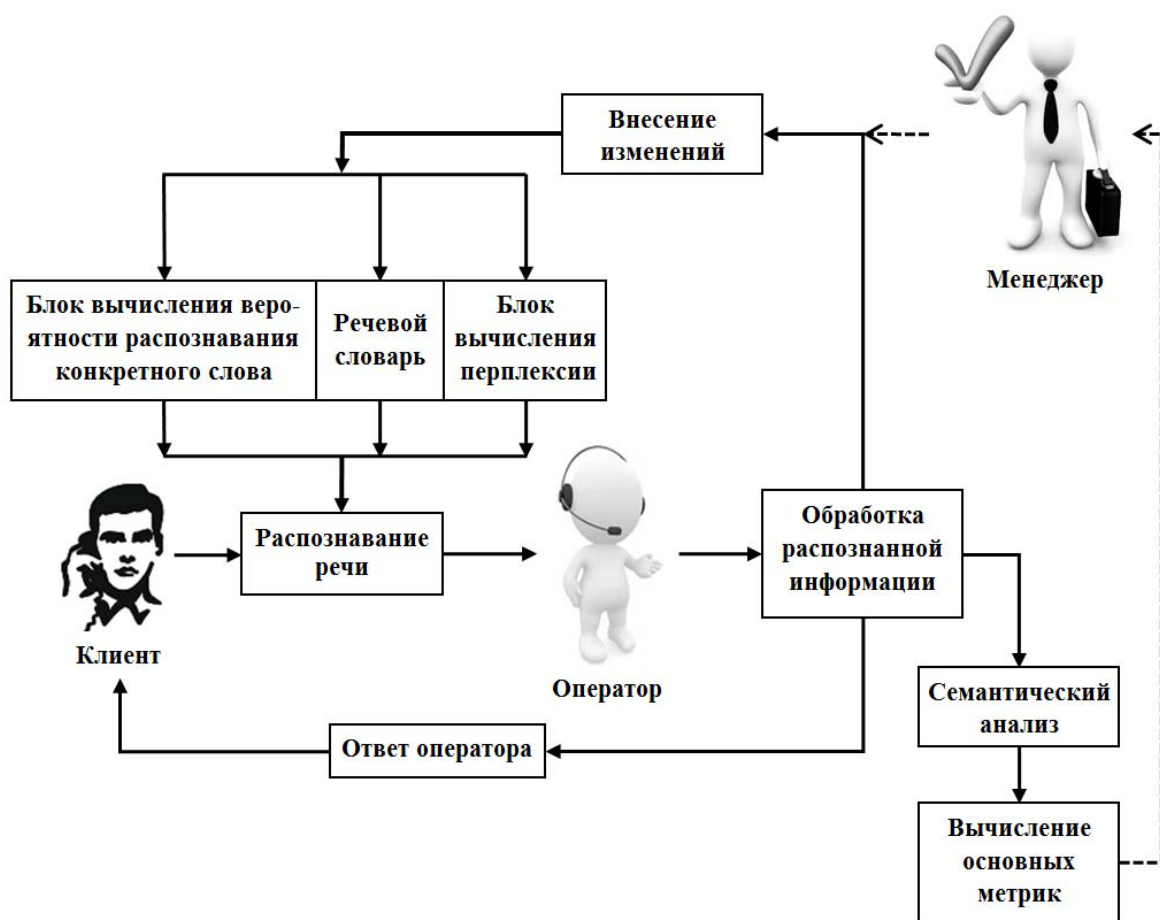


Рис. 1. – Предлагаемая система распознавания речи

В рамках проведенных исследований по предлагаемой системе с использованием нескольких предметных областей, не связанных между собой [5], было определено, что для русского языка словарь размером 2500 слов является достаточным для покрытия более 95% информации конкретизированной (усреднённой) предметной области. Это означает, что не менее чем 95% голосового сигнала будет визуализировано с точностью,

которая характерна для системы распознавания. Но учитывая, что русский язык является флективным языком, данный размер может быть увеличен в пять-семь раз для покрытия всего существующего в языке числа словоформ.

Дополнительным подтверждением экспериментально полученного значения 2500 слов как достаточного является графическое решение задачи оптимизации точности *Acc* (англ. *Accuracy*) и скорости *Ran* (англ. *Range*) распознавания по отношению к размеру словаря *Voc* (англ. *Vocabulary*) с помощью суммирующей функции *Summ*. В результате проведённых исследований с использованием комбинации пяти вероятных словарей телефонных служб поддержки было получено значение в размере около 2600 слов (рис. 2). Значение в пять секунд было выбрано как пороговое для задержки распознавания, поскольку превышение данного порога будет оказывать влияние на работу оператора. Таким образом, метрика *RTF* (англ. *Real time factor*) для фразы длиной 20 секунд не превысит значения 0,25.

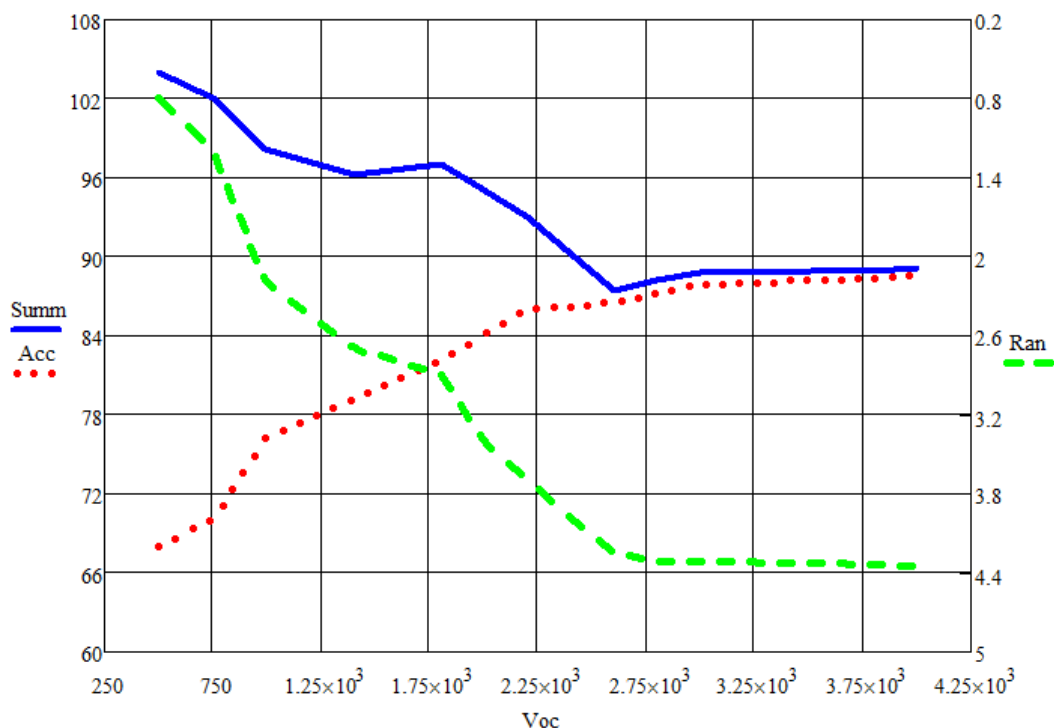


Рис. 2. – Решение задачи оптимизации размера словаря

В итоге дополнительная возможность модификации речевого словаря непосредственно во время работы системы распознавания обеспечивает возможность непрерывного повышения точности распознавания речи. Наиболее очевидно данное преимущество проявляется в работе телефонных служб поддержки, потому что каждая из них имеет строго конкретизированную предметную область [4]. Под конкретизированной (усреднённой) предметной областью в работе понимается малая область знаний, в которой затрагивается лишь несколько конкретных вопросов. В качестве примеров можно привести: использование лучевой терапии при лечении онкологических заболеваний, операции с банковскими картами, логистические маршруты, бронирование билетов в турфирме.

Безусловно, существуют такие предметные области, в которых словарь будет больше в несколько раз. Такой словарь важен для служб такси крупных городов или телефонных центров аптек. Но значение речевого словаря в 2500 слов позволит эффективно внедрить систему распознавания, провести первичную подстройку и лишь затем расширять речевой словарь.

Перплексия и ограничение словаря

Необходимо отметить ещё одну важную особенность предлагаемой системы. Автором предложено учитывать связность слов, или так называемую перплексию. Перплексия является количественной мерой оценки среднего числа слов, которые могут следовать после какой-либо N -граммы, где N -граммная модель – статистическая модель, которая позволяет моделировать язык. Следовательно, связность учитывает число слов, которые логически могут встретиться после данного слова. Именно связность повышает точность распознавания речи, поскольку в большинстве случаев вопросы, задаваемые клиентами контакт-центров, имеют достаточно четкий синтаксис, что характерно прежде всего для английского языка. Также

перплексия помогает учитывать психо-эмоциональное состояние звонящего клиента, ведь его речь будет содержать ещё меньшее число слов [6,7,8].

Для этого выделен специальный блок вычисления перплексии (рис. 1), который используется во время распознавания. В этот блок с помощью блока обратной связи регулярно вносятся изменения на основе результатов обработки распознаваемой информации. Эти изменения касаются не только добавления новых слов, но и ранжирования слов, которые могут встречаться после какого-либо другого слова. Такая особенность обеспечивает как общее увеличение точности распознавания, так и учёт региональных и сезонных колебаний запросов от клиентов.

Перплексия вычислялась по формуле [9]:

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

где N – число состояний модели q , а x_i – состояние модели. Меньшее значение перплексии характерно для ситуаций, в которых модель лучше обеспечивает предсказание следующего слова.

Необходимо отметить, что для телефонных служб поддержки клиентов важным является не только распознавание (визуализация) звонка, но и дальнейший анализ поступившей информации. Исходя из этого, необходимо сделать акцент не только на формальной стороне каждого звонка, но и на содержательной [10]. Поэтому в предлагаемой системе существует возможность анализа менеджером ключевых метрик и ключевых слов [11], что позволит оптимизировать размер речевого словаря за счёт уточнения наиболее частых слов, добавления наиболее значимых и даже удаления редких слов.

Заключение

В работе рассматриваются факторы, влияющие на размер речевого словаря. Показано, что для телефонной службы поддержки, имеющей

конкретную усреднённую предметную область, нижняя граница размера речевого словаря русской речи имеет размер в 2500 слов. Данное значение подтверждается как экспериментально, так и графически. Использование ограниченного речевого словаря позволит сократить время на распознавание речи клиента, повысить удовлетворённость клиентов и производительность телефонных служб поддержки, а также снизить затраты на их работу.

Литература

1. Хижинский Д. Распознавание речи взорвет рынок call-центров // cnews.com: сетевой журн. 22.05.2007. URL: cnews.ru/reviews/index.shtml?2007/05/22/251191 (дата обращения: 22.07.2014).
2. How is speech recognition applied in call centers? URL: glocal.com/voice-recognition/speech-recognition-in-call-centers.html (access date: 30.02.2015).
3. Бальчюнене Н.И. К вопросу разработки профессиональных русско-финских (финско-русских) словарей // Инженерный вестник Дона. – 2014. – № 1. URL: ivdon.ru/ru/magazine/archive/n1y2014/2257.
4. Балакшин П.В., Петров Г.Ю. Некоторые аспекты исследования систем распознавания речи в телефонных службах поддержки // Научно-технический вестник информационных технологий, механики и оптики. – СПб: Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики (СПбНИУ ИТМО). – 2012. – Вып. 1. – № 77. – Компьютерные системы и информационные технологии. – С. 71-76. – ISSN 2226-1494.
5. Балакшин П.В., Тропченко А.Ю. Выбор размера словаря при реализации тестовой системы распознавания речи // Интеллектуальные и информационные системы: Материалы Всероссийской научно-технической конференции / Тульский государственный университет. – Тула, 2009. – С. 51-54.



6. Сидоров К.В., Ребрун И.А., Кожевников Д.Д., Сobotницкий И.С. Диагностика психофизиологического и эмоционального состояния человека-оператора // Инженерный вестник Дона. – 2012. – № 4 (часть 2). URL: ivdon.ru/ru/magazine/archive/n4p2y2012/1480.

7. Марьев А.А. О возможности повышения эффективности автоматического интонационного анализа речи // Инженерный вестник Дона. – 2012. – № 3 (часть 2). URL: ivdon.ru/ru/magazine/archive/n3y2012/898.

8. Потапова Р.К., Потапов В.В. Модификация речевого сигнала как следствие наличия эмоциональных состояний «страх»/«тревожность» // Речевые технологии. – 2012. – № 1. – С. 52-60.

9. Call Centre Monitoring / callcentrehelper.com: сетевой журн. 2010. URL: callcentrehelper.com/call_centre_monitoring.htm (access date: 03.06.2015).

10. Бабин Д.Н., Мазуренко И.Л., Холоденко А.Б. О перспективах создания системы автоматического распознавания слитной устной русской речи // Интеллектуальные системы. – М. – 2004. – Том 8, вып. 1-4. – С. 45-70.

11. Гусев М.Н., Дегтярев В.М. Выделение ключевых слов // Речевые технологии. – 2012. – № 1. – С. 15-21.

References

1. Hizhinskij D. cnews.com: setevoj zhurn. 22.05.2007. URL: cnews.ru/reviews/index.shtml?2007/05/22/251191.

2. How is speech recognition applied in call centers? URL: glocal.com/voice-recognition/speech-recognition-in-call-centers.html.

3. Bal'chjunene N.I. Inzhenernyj vestnik Dona (Rus), 2014. № 1. URL: ivdon.ru/ru/magazine/archive/n1y2014/2257.

4. Balakshin P.V., Petrov G.Ju. Nauchno-tehnicheskij vestnik informacionnyh tehnologij, mehaniki i optiki. Saint-Petersburg, 2012. Issue 1. № 77. Pp. 71-76. ISSN 2226-1494.



5. Balakshin P.V., Tropchenko A.Ju. Materialy Vserossijskoj nauchno-tehnicheskoy konferencii «Intellectual'nye i informacionnye sistemy» [Proc. All-Russian scientific-technical conf. «Intellectual and Information Technologies»]. Tula, 2009. Pp. 51-54.
6. Sidorov K.V., Rebrun I.A., Kozhevnikov D.D., Sobotnickij I.S. Inzhenernyj vestnik Dona (Rus), 2012. № 4 (chast' 2). URL: ivdon.ru/ru/magazine/archive/n4p2y2012/1480.
7. Mar'ev A.A. Inzhenernyj vestnik Dona (Rus), 2012. № 3 (chast' 2). URL: ivdon.ru/ru/magazine/archive/n3y2012/898.
8. Potapova R.K., Potapov V.V. Rechevye tehnologii, 2012. № 1. Pp. 52-60.
9. Call Centre Monitoring. callcentrehelper.com. 2010. URL: callcentrehelper.com/call_centre_monitoring.htm.
10. Babin D.N., Mazurenko I.L., Holodenko A.B. Intellectual'nye sistemy. Moscow. 2004. Vol. 8, Issue 1-4. Pp. 45-70.
11. Gusev M.N., Degtjarev V.M. Rechevye tehnologii, 2012. № 1. Pp. 15-21.