Выделение ключевых слов

Гусев М.Н., *кандидат технических наук, ООО «Вокатив».*

Дегтярёв В.М., *доктор технических наук, профессор СПбГУТ*

В статье описываются преимущества использования систем поиска ключевых слов для обеспечения безопасности. Рассматриваются различные подходы к построению систем поиска ключевых слов, анализируются их достоинства и недостатки, выбирается оптимальный вариант. Описывается общий алгоритм работы системы, вводятся основные критерии оценки её качества. Приводятся результаты тестирования разработанного решения.

• *распознавание речи* • *поиск ключевых слов*.

The article presents advantages of keyword spotting systems being used for security maintenance. Various approaches to construction of keyword spotting systems are considered, their «pros» and «cons» are analyzed, finally the optimum variant being chosen. The general algorithm is outined and the basic criteria of its quality estimation are given. Test results of the developed solution are presented.

• *speech recognition* • *key-word spotting*.

Введение

Речь — это идеальное, доступное средство передачи информации, первичный языковой навык и неотъемлемый инструмент общения. Для человека говорить так же естественно, как есть или спать [1]. Чтобы лучше понимать, что происходит в компании, службе безопасности, полезно знать, о чём говорят сотрудники, клиенты и контрагенты.

Отслушка переговоров давно входит в набор средств обеспечения безопасности. Обычно ведётся точечный контроль сотрудников из «группы риска». Регулярная работа создаёт

эффективную среду противодействия рискам. Однако у данного подхода есть ряд недостатков:

1) неполнота охвата и недостаточная эффективность;
2) высокая вероятность пропустить «подозрительный» разговор;
3) отслушка записей переговоров незаконна, так как для неё требуется соответствующее решение суда.

Автоматизация отслушки позволяет сделать охват полным и во много раз повысить оперативность получения данных. Кроме того, решается вопрос с законностью, поскольку на отслушке работает не человек, а машина.

Возможные подходы к построению системы

Поиск ключевых слов работает очень просто: на вход системы подаются записи переговоров и список искомых ключевых слов и фраз (КС), на выходе получается список подозрительных разговоров (см. рис.1).

Рис.1. Схема работы системы

За кажущейся простотой работы системы скрывается реализация сложных математических и лингвистических алгоритмов. Возможны различные подходы к решению задачи поиска КС, обладающие достоинствами и недостатками:

- KWS[1] на основе динамического программирования;
- KWS на фоновой сети;
  - на монофонной сети;
  - на трифонной сети;
- KWS на основе ASR;
  - по словным латтисам;
  - по фоновым латтисам;
- KWS на моделях ключевых слов (КС).

При создании системы KWS на основе принципов динамического программирования КС произносится несколько раз несколькими дикторами. По произнесённым словам строится шаблон слова, который ищется в потоке речи. Это неудобно, так как для каждого искомого слова требуется создавать свой шаблон. Создание нового шаблона и смена списка искомых КС оказывается трудоёмкой и финансово затратной операцией.

Часто для поиска КС используются те же модели, что и для распознавания слитной речи. Сначала обучаются модели отдельных звуков. Затем по моделям звуков строятся фоновые (или фонетические) сети, или модели КС.

[1] От английского key-word spotting – поиск ключевых слов.

На рис. 2 представлен пример структуры данных, используемой в KWS системе, основанной на фоновой сети.

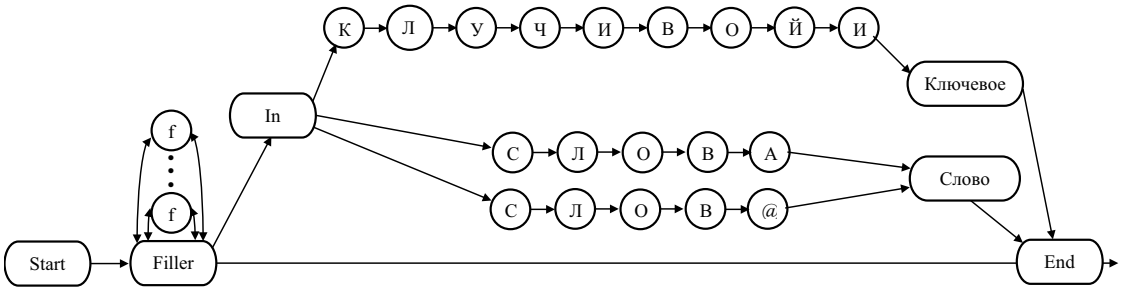

Рис. 2. Пример структуры фоновой сети

В зависимости от используемых моделей звуков фоновые сети разделяются на монофонные и трифонные. В первом случае используются звуковые модели, не учитывающие звуковой контекст и дающие менее точное описание речевого сигнала. Во втором случае точность описания КС значительно повышается, но снижается скорость обработки звука, поскольку увеличивается количество задействованных звуковых моделей. Конечно, возможно применение методов оптимизации фонетической сети, но возможности ускорения упираются в количество и структуру искомых КС.

Во втором, используемом нами, варианте из моделей звуков собираются модели КС. Кроме моделей КС, строятся модели заполнения, описывающие шумы и неречевые сигналы. Также создаются модели речевого мусора или модели усреднённого речевого потока, которые описывают все остальные слова, не являющиеся искомыми.

Для каждого КС строятся свои модели заполнения и усреднённого речевого потока, что позволяет оптимизировать их структуру и увеличить качество поиска. Пример структуры данных, используемой в KWS системе на основе моделей КС, приведён на рис. 3.

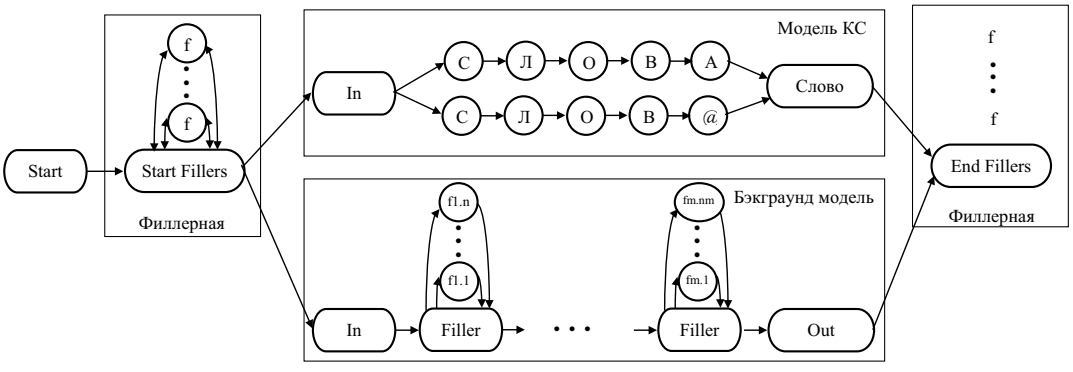

Рис. 3. Пример структуры данных в KWS системе на моделях КС

Ещё один подход к поиску КС основан на распознавании речи в чистом виде. В результате работы системы распознавания речи формируется латтис — направленный связный граф, некоторая сеть слов, содержащая слова кандидаты на распознавание, связи между ними и вероятности переходов. Наилучший вариант пути в графе (имеющий наибольшую вероятность) присутствует в графе, однако графом описывается некоторое количество конкурирующих гипотез. Латтис, в узлах которого находятся слова, называется словным латтисом.

На основании словных латтисов, формируемых системой распознавания с большим словарём, выполняется индексация звукового массива. Далее поиск КС и фраз выполняется по полученным словным латтисам. Достоинством такой системы является высокая скорость поиска КС в индексированных звуковых данных. Проблема такой системы — в принципиальной невозможности нахождения слова, отсутствующего в словаре системы распознавания. Кроме того, такие системы поиска КС оказываются сильно завязанными на качество работы систем распознавания.

Альтернатива словного латтиса — фонемный латтис, в узлах которого находятся не слова, а отдельные звуки речи. Система распознавания аналогично словному латтису строит фонемный латтис, по которому и выполняется поиск КС.

Преимуществом фонемного латтиса является возможность поиска любых КС, так как в системе распознавания не используется словарь и отсутствует привязка к словам. Сложность такой системы заключается в том, что фонемный латтис оказывается широким, качество фонемного распознавания — низким, а пространство поиска — велико.

Основные элементы разработанной системы

В результате анализа достоинств и недостатков различных принципов построения KWS систем было решено создавать систему на основе моделей КС. Основу разрабатываемой системы поиска ключевых слов составляют следующие модули:

- база HMM-моделей звуков речи;
- автоматический транскриптор ключевых слов;
- звуковой препроцессор, выполняющий предварительную обработку звукозаписей и преобразование звука в параметры;
- формирователь альтернативных моделей (моделей усреднённого речевого потока и моделей заполнения);
- декодер параметризированного звука.

Система работает по следующему алгоритму:

- транскриптор формирует возможные варианты произнесения искомых ключевых слов и фраз;
- для всех ключевых слов формируются альтернативные модели;
- полученные структуры данных объединяются в общую сеть поиска — рабочую структуру системы распознавания;
- звуковой поток обрабатывается препроцессором и переводится в пространство признаков;
- звуковой поток разделяется на окна. Каждое окно подаётся на вход декодера. Длины окон и параметры перекрытия определяются исходя из звукового состава искомых слов;
- декодер анализирует параметризированный речевой поток и принимает решение о наличии или отсутствии ключевых слов. Декодирование выполняется с помощью модифицированного алгоритма пересылки маркера [2];
- получаемые результаты распознавания привязываются к звуковому потоку и сохраняются в специальном индексном файле.

Оптимизация структур данных

Для увеличения скорости поиска возможна оптимизация рабочих структур данных распознавателя [3]. Оптимизация заключается в объединении одинаковых цепочек фонов, в рамках вариантов различных транскрипций КС.

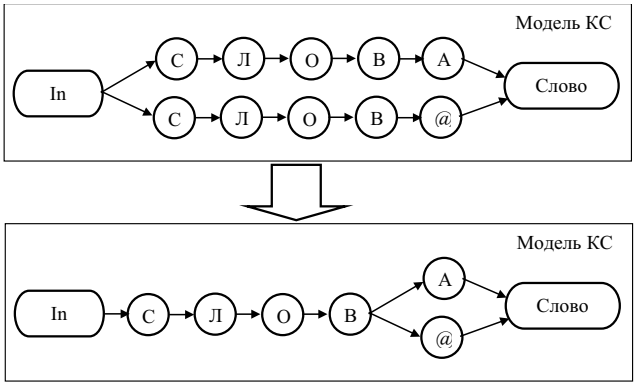

Рис. 4. Пример оптимизации модели КС

Оптимизация может быть продемонстрирована на примере структуры данных, представленной на рис. 3. В результате объединения одинаковых цепочек фонов модель КС примет вид, как показано на рис. 4. Может показаться, что с учётом общего количества звуковых моделей, используемых в моделях заполнения и усреднённого речевого потока, получаемая оптимизация не велика. На самом деле, она позволяет сократить от 5 до 20% всех звуковых моделей. Всё зависит от сложности общей структуры и вариативности произношения искомого КС, его длины и звукового состава. Чем длиннее КС и чем выше возможная вариативность произнесений, тем больше оптимизация.

Критерии оценки качества системы

Одна из важнейших характеристик системы поиска ключевых слов — точность. Под точностью понимается пара значений: DR (Detection Rate) и FA (False Alarm).

Значение DR определяет процент правильно обнаруженных слов и рассчитывается по формуле:

$DR = 100 * N_{found} / N_{all}$, где:

N_{found} — количество правильно найденных реализаций КС в тестовых данных;

N_{all} — общее количество реализаций КС в тестовых данных.

Значение FA определяет количество ложных срабатываний в час и рассчитывается по формуле:

$FA = (A_{all} — N_{found}) / Hrs$, где:

A_{all} — общее количество всех найденных слов в тестовых данных;

Hrs — длительность звучания тестовых данных в часах.

Результаты тестирования

Тестирование системы проводилось на звуковых файлах общей длительностью звучания чуть больше часа (1 час 2 минуты 37 секунд). Звонки были выполнены с городских телефонных аппаратов. В тестировании приняли участие десять дикторов (шесть мужчин и четыре женщины).

В результате тестирования различных режимов построения моделей заполнения и моделей усреднённой речи были получены значения точности работы системы, представленные в таблице. Тестировались следующие режимы:

«Тюнингованные» — в данном режиме параметры поиска и альтернативные модели подбираются вручную для каждого КС на множестве обучающих звуковых данных;

«Смарт-авто» — система автоматически формирует параметры поиска и альтернативные модели для каждого слова исходя из их фонетического состава КС и известных параметров фонем;

«Стандарт» — используется фиксированный набор параметров поиска и альтернативных моделей;

«Короткий» — режим, аналогичный режиму «Стандарт», но с меньшей вариативностью моделей заполнения.

Показатели качества работы системы

Ключевое слово	Тюнингованные		Смарт-авто		Стандарт		Короткий	
	DR	FA2	DR	FA2	DR	FA2	DR	FA2
Кодовое слово	100,00	38,49	100,00	35,00	100,00	400,97	96,00	1350,48
Кредит	96,00	170,00	70,00	32,00	93,00	362,00	96,00	3570,27
Задолженность	100,00	48,11	83,00	6,41	100,00	131,52	83,00	1376,14
Номер карты	83,00	38,49	63,00	0,00	96,00	109,06	76,00	1661,63
Конфиденциально	95,00	79,57	61,00	19,66	88,00	230,67	74,00	982,59
Вакансии	95,00	84,04	77,00	4,47	91,00	167,00	84,00	1957,15
	94,83	76,45	75,67	16,26	94,67	233,54	84,83	1816,38

Видно, что показатели качества работы системы во многом зависят от режима работы системы и от самого искомого КС. Способы построения альтернативных моделей требуют дальнейших исследований и разработки.

Литература

1. *Смирнов В., Ермилов С*. Технология распознавания речи на службе корпоративных интересов // «Директор по безопасности». 2010. № 11. С. 27–37.

2. *Young S.J., Russell N.H., Thornton J.H.S*. Token Passing: a Conceptual Model for Connected Speech Recognition Systems // CUED Technical Report F INFENG/TR38, Cambridge University, 1989. (ftp from svr-ftp.eng.cam.ac.uk)

3. *Гусев М.Н., Дегтярев В.М*. Увеличение производительности системы распознавания речи. «Вопросы радиоэлектроники» (Серия Общетехническая), 2010. Вып. 2. С. 115–126.

Сведения об авторах

Дегтярёв Владимир Михайлович —
заведующий кафедрой Санкт-Петербургского государственного университета телекоммуникаций им. проф. М.А. Бонч-Бруевича, д. т. н., профессор, академик Международной академии информатизации, секция «Аудио-, видеоинформации».

Гусев Михаил Николаевич —
Главный инженер-программист ООО «Вокатив», к. т. н. Научные интересы — обработка речевых сигналов и распознавание речи.

Публикации

1. *Дегтярёв В.М., Гусев М.Н*. Развитие систем распознавания речи // Труды СПбГТУ «Вычислительная техника, автоматика, радиоэлектроника»/ СПбГТУ. СПб, 2006. № 499. С. 119–124.

2. *Valentin Smirnov, Mikhail Gusev*. Objective method of speech signal quality estimation // Proceedings of the 11-th International Conference «Speech and Computer» SPECOM'2006. St.Petersburg, Anatolya Publishers, 2006. Pp. 242–244.

3. *Гусев М.Н., Смирнов В.А., Дегтярев В.М*. Компьютерная статистическая модель русского языка // Труды учебных заведений связи / СПбГУТ. СПб, 2006. № 174. С. 129–135.

4. Патент РФ № 2296377. Способ анализа и синтеза речи. Гусев М.Н., Дегтярёв В.М., Ситников В.В. Официальный Бюллетень Федеральной службы по интеллектуальной собственности, патентам и товарным знакам. Изобретения. Полезные модели, 27.03.2007, № 9(2), 2007.

5. Патент РФ № 2312405. Способ осуществления машинной оценки качества звуковых сигналов, Гусев М.Н., Дегтярёв В.М., Жарков И.В. Официальный Бюллетень Федеральной службы по интеллектуальной собственности, патентам и товарным знакам. Изобретения. Полезные модели, 10.12.2007, № 34(2), 2007.

6. *Гусев М.Н., Дегтярёв В.М., Семёнов Н.Н*. Оптимизация системы распознавания речи с учётом особенностей артикуляции. // Труды учебных заведений связи / СПбГУТ. СПб, 2007. № 177. С. 20–24.

7. *Bolotova Olga, Gusev Michael, Smirnov Valentin*. Speech Recognition System for the Russian Speech // Proceedings of the 12-th International Conference «Speech and Computer» SPECOM'2007. Moscow, 2007. V.II. Pp.475–480.

8. *Гусев М.Н., Дегтярев В.М*. Расчёт и измерение качества речевых сигналов. Санкт-Петербург: «Геликон Плюс», 2008. 276 с.

9. *Смирнов В., Ермилов С*. Технология распознавания речи на службе корпоративных интересов // «Директор по безопасности», 2010. № 11. С. 27–37.

10. *Гусев М.Н., Дегтярёв В.М*. Моделирование длительности звуков в системе распознавания речи. «Вопросы радиоэлектроники» (Серия Общетехническая), 2010. Вып. 2. С. 106–115.