Комплексная система верификации ключевых слов на основе метода опорных векторов

А.М. Сорока,
БГУ, Минск, Беларусь

Одной из основных сложностей, влияющих на точность методов поиска ключевых слов (ПКС), остаётся проблема декодирования акустически схожих пар спутывания, точность решения которой в значительной степени влияет на общую эффективность системы [1]. Это свидетельствует о необходимости дополнительного анализа пар спутывания, представляющих конкурирующие версии. Кроме этого, при использовании оценок апостериорных вероятностей для принятия решений в системах ПКС высокая точность поиска сопряжена с высоким уровнем ложных тревог, для уменьшения которого необходимо использовать специализированные алгоритмы верификации. В статье рассмотрен алгоритм верификации потенциальных ключевых слов на основе метода опорных векторов (МОВ).

Основные методы верификации речевых данных

В системе поиска ключевых слов неизбежны ошибки, связанные в первую очередь с особенностями и функциональным состоянием диктора, шумами окружающей среды, помехами и т.д. Использование решётки слогов позволяет проанализировать максимальное количество вариантов последовательностей, содержащих искомые ключевые слова, и тем самым максимизировать вероятность правильного обнаружения [2]. Однако такие процедуры приводят к увеличению количества ложных тревог. Использование алгоритмов верификации найденных ключевых слов позволит в значительной степени снизить чувствительность системы к словам, не входящим в словарь, т.е. уменьшить вероятность ложной тревоги.

Существует три наиболее распространённых метода верификации:

1) Метод, основанный на использовании эталонов. Метод применяется только в условиях, когда ключевые и «неключевые» слова фиксированы на этапе создания систе-мы. Любое изменение словаря и условий использования приводит к необходимости

полного переобучения системы, что неприемлемо для большинства практических приложений.

2) Статистический метод. Эффективность метода сильно зависит от выбранного признака верификации и классификатора. Наиболее распространённые признаки — соотношение правдоподобия, апостериорная вероятность слова, акустическая вероятность нормализации; параметры лингвистической модели, акустическая устойчивость, контекст ключевых слов. В качестве классификаторов могут использоваться байесовские классификаторы, нейронные сети и СММ.

3) Ассоциативный метод на основе статистических данных и шаблонов правил. Пример такого подхода — алгоритм TBL, предложенный Э. Брилем [3].

Наиболее распространённый и эффективный — подход верификации данных с использованием статистических моделей. Главная причина успешного использования СММ — существование итеративных методов, гарантирующих сходимость оценки параметров. Однако поскольку СММ генерируется на основе традиционной статистической теории распределения вероятностей, СММ может корректно описывать модель только в том случае, если есть достаточное количество обучающих данных. Кроме этого, СММ позволяет точно идентифицировать данные разных типов только тогда, когда в пространстве выборок область распределения данных различных типов или не перекрывается, или перекрывается незначительно.

Искусственные нейронные сети представляют интересный и важный класс классификаторов, успешно использованный для анализа и распознавания речи. Применение ИНС позволяет преодолеть многие недостатки, свойственные СММ, однако использование ИНС для задач верификации имеет ряд недостатков. Наиболее значимые недостатки — сложность определения оптимальной топологии модели, медленная сходимость во время обучения и тенденция к чрезмерной адаптации данных.

С точки зрения верификации ключевых слов в динамических условиях наиболее перспективным выглядит классификатор на основе МОВ. Даже при небольшом количестве обучающих данных, при помощи МОВ в пространстве признаков можно найти самую оптимальную гиперплоскость для создания классификатора, что обуславливает её широкое применение на настоящий момент. МОВ также эффективно применяется для задач распознавания речи, изображений и др. Впервые для обработки речи МОВ был использован А. Ганапавираем в 1998 г. В его работе в качестве предварительного препроцессора для сегментации по фонемам была использована СММ, затем производилось принудительное выравнивание по фиксированной длине и масштабу (3:4:3), а для последующего распознавания речи была использована гибридная модель на основе СММ и МОВ [4]. Результаты применения такой системы показали, что она эффективнее, чем изолированная модель СММ.

Наиболее удачны бинарные МОВ, способные разделять данные сложной и схожей конфигурации на два класса, что делает их весьма эффективными для решения задач верификации и идентификации близкорасположенных данных. В связи с этим в данной статье предлагается использовать МОВ для верификации ключевых слов в системах ПКС.

Верификация ключевых слов с использованием меры достоверности на интервале

Представим речевой сигнал на входе системы поиска ключевых слов в виде последовательности наблюдений $O = \{O_1, O_2, ..., O_T\}$. Поскольку ключевые слова уже найдены, то известно начало и конец каждого слова, и параметры их СММ могут быть оценены на основе решётки. Обозначим такую СММ ключевого слова, как Δ, и определим меру достоверности как количественную величину совпадения O и Δ. Другими словами, мера достоверности определяется вероятностью генерации последовательности O на основе модели Δ.

Для моделирования речевых сигналов на основе СММ компоненты вектора признаков предполагаются независимыми, и для каждого состояния плотность распределения вероятностей наблюдений можно представить смесью гауссовых распределений, задающей вероятность вектора наблюдений O_t в состоянии j как:

$$b_j(o_t) = \prod_{d=1}^{D}\left[\sum_{m=1}^{M_s} c_{jdm} N\left(o_{dt}; \mu_{jdm}; \sigma_{jdm}\right)\right],$$

где D — размерность вектора признаков речевого сигнала; M_s — размерность ГС для компоненты d; c_{jdm}, σ_{jdm}, μ_{jdm} — соответственно весовое значение, среднее значение и дисперсия для m-й компоненты смеси, предположительно описывающуюся нормальным распределением:

$$N(o_{dt}; \mu_{jdm}; \sigma_{jdm}) = \frac{1}{\sqrt{2\pi}\sigma_{jdm}} e^{-(o_{dt}-\mu_{jdm})^2 \Big/ 2\sigma_{jdm}^2}$$

Пусть j-е состояние модели Δ определяется участком последовательности $O^k = \{O_k, O_{k+1}, ..., O_K\}$, тогда значение меры достоверности CM_j можно определить следующим образом:

$$CM_j = \frac{\sum_{t=k}^{K}\sum_{m=1}^{M_s}\sum_{d=1}^{D} F\left(o_{dt}; \mu_{jdm}, \sigma_{jdm}\right)}{(K-k)M_d S},$$

где

$$F(o_{ti}; \mu_{jdm}, \sigma_{jdm}) = \begin{cases} 1, & (o_{dt}-\mu_{jdm}) \in [\mu_{jdm} - k\sigma_{jdm}, \mu_{jdm} + k\sigma_{jdm}] \\ 0, & \check{c}\acute{i}\acute{r}\div\acute{l} \end{cases},$$

где k — управляющий интервалом достоверности параметр. Определим меру достоверности для каждого ключевого слова Δ путём нормализации значений для каждого состояния:

$$CM_1 = \frac{1}{N}\sum_{j=1}^{N} CM_j,$$ где N — количество состояний СММ.

Верификация ключевого слова может происходить путём сравнения полученной меры достоверности с некоторым пороговым значением, как правило, выбранным эмпирически.

Верификация ключевых слов на основе нормализации длительности состояния

Один из недостатков меры достоверности, представленной выше, — отсутствие её нормализации на длину состояния СММ, вследствие этого возможны ситуации, когда

состояние с малой длительностью затеняет результаты для более длительной последовательности наблюдений.

Пусть есть СММ ключевого слова $\lambda = \{N_1, \pi, A, B\}$, где N — число состояний СММ $S = \{S_1, S_2, \ldots S_N\}$, π — матрица начальных вероятностей, A и B — матрицы вероятностей переходов и наблюдений соответственно. Обозначим время входа и выхода системы из состояния i как $b[i]$ и $e[i]$ соответственно, тогда нормализованную меру достоверности можно определить следующим образом:

$$CM_2 = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{1}{e[i]-b[i]+1}\sum_{t=b[i]}^{e[i]}\log p(o_t \mid s_i)\right] = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{1}{e[i]-b[i]+1}\sum_{t=b[i]}^{e[i]}\log b_i(o_t)\right]$$

Согласно формуле Байеса заменив $\log p(o_t \mid s_i)$ на $\log p(s_i \mid o_t)$ получим ещё одно выражение для меры достоверности:

$$CM_3 = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{1}{e[i]-b[i]+1}\sum_{t=b[i]}^{e[i]}\log p(s_i \mid o_t)\right]$$

$$p(s_i \mid o_t) = \frac{p(o_t \mid s_i)p(s_i)}{\sum_{j=1}^{N} p(o_t \mid s_j)p(s_j)} = \frac{b_i(o_t)p(s_i)}{\sum_{j=1}^{N} b_j(o_t)p(s_j)}$$

Представленные меры достоверности представляют собой среднее значение акустической вероятности в рамках СММ [5].

Верификация ключевых слов на основе динамического рейтинга

Описанные выше методы подтверждения ключевых слов в той или иной степени используют модели фонем. Они просты и эффективны, в значительной степени позволяют снизить вероятность ложной тревоги. Рассмотрим ещё один метод верификации ключевых слов, основанный на использовании апостериорной сети и динамического рейтинга [6].

В результате работы декодирующего алгоритма Витерби текущему вектору наблюдений o_t входящего речевого сигнала и каждой допустимой в данный момент модели слова Δ_j ставится в соответствие последовательность состояний модели $S(\Delta_j, o_t)$. Определим меру соответствия (характеристическое значение) $L_j(o_t)$ последовательности состояний $S(\Delta_j, o_t)$ модели Δ_j в момент времени t следующим образом:

$$L_j(o_t) = \ln P\left(o_t \mid S\left(\Lambda_j, o_t\right)\right), \quad j = 1, 2, \ldots, N(o_t) \quad (1)$$

где $N(o_t)$ — число допустимых моделей в момент времени t. Отсортируем набор характеристических значений по убыванию для всех моделей:

$$L_\Lambda(o_t) > L_{j_2}(o_t) > \ldots > L_{j_k}(o_t) > \ldots > L_{j_{N(o_t)}}(o_t)$$

Пусть характеристическое значение $L_{Kw}(o_t) = L_{jk}(o_t)$ для модели ключевого слова Δ_{Kw} занимает в отсортированной последовательности позицию k, тогда динамический рейтинг на O_t-м фрейме можно определить как $k/N(o_t)$:

$$Q(o_t \mid \Lambda_{Kw}) = \frac{\sum_{k=1}^{N(o_t)} G(L_k(o_t) - L_{Kw}(o_t))}{N(o_t)},$$

где

$$G(L_k(o_t) - L_{Kw}(o_t)) = \begin{cases} 0, & L_k(o_t) \le L_{Kw}(o_t) \\ 1, & \check{c}\acute{i}\,\acute{r} \div \acute{I} \end{cases}$$

Введём меру достоверности на основе динамического рейтинга как обобщённое характеристическое значение в рамках всей анализируемой длительности сигнала:

$$CM_4(O \mid \Lambda_{Kw}) = \frac{1}{T}\sum_{t=1}^{T} Q(o_t \mid \Lambda_{Kw})$$

Этот метод пороговый, при этом для всех ключевых слов может быть установлено одинаковое пороговое значение.

Вычисление динамического рейтинга для ключевых слов не приводит к существенному увеличению вычислительной сложности системы по сравнению с алгоритмами на основе акустической меры достоверности. В процессе декодирования Витерби значение вероятностей для выражения (1) уже рассчитаны, поэтому вычисление $Q(o_t)$ и последующая нормализация приводят к дополнительным DT сложениям и вычитаниям, а также $T+1$ делению. Однако этот алгоритм обладает рядом существенных преимуществ. Во-первых, эффективность верификации ключевых слов на основе динамического рейтинга имеет более стабильный характер, особенно в условиях шума. Это обусловлено тем, что имеющийся в речевых данных шум влияет только на абсолютные значения акустических вероятностей и не оказывает практически никакого влияния на расположение характеристических значений при сортировке. Во-вторых, динамический рейтинг рассчитывается исключительно на основании значений акустической вероятности, поэтому изменение словаря ключевых слов не оказывает никакого влияния на работоспособность алгоритма. В-третьих, как уже упоминалось выше, для всех ключевых слов может быть установлено одинаковое пороговое значение, и кроме этого, наличие шума в исходных данных не влияет на абсолютное значение порога.

Верификация ключевых слов на основе машины на опорных векторах

Все методы верификации ключевых слов, основанные на моделях фонем и мерах достоверности, обладают общими недостатками. Во-первых, эти методы пороговые, процедура определения порога имеет, как правило, эмпирический характер и сопряжена с проведением многочисленных экспериментов. Во-вторых, алгоритм верификации ключевых слов на основе единичной меры доверительности, как и любая пороговая система, обладает ограничениями по улучшению его характеристик. При сохранении высокой точности распознавания вероятность ложной тревоги может быть уменьшена только до какого-то фиксированного значения, характерного для конкретной меры доверительности.

Для решения проблемы можно использовать алгоритмы верификации ключевых слов на основе нескольких мер достоверности одновременно. Существует много методов объединения нескольких мер доверительности, например, логическое соединение, линейный дискриминант Фишера, метод нейронной сети и т.д. Предположим, что каждому ключевому слову поставлено в соответствие u мер доверительности $CM_1, CM_2, \ldots CM_u$, каждой из которых соответствует своё пороговое значение $TH_1, TH_2, \ldots TH_u$. В простейшем случае значения мер достоверности $cm_1, cm_2, \ldots cm_u$ сравниваются с соответствующими порогами, и если $cm_1 > TH_1, cm_2 > TH_2, \ldots cm_u > TH_u$, то это слово определяется как правильно распознанное, иначе принимается решение об ошибке распознавания. Введение

различных весовых коэффициентов позволяет гибко настраивать систему верификации, однако улучшение характеристик такой системы тоже ограничено, поскольку данный метод также пороговый. Для преодоления этих недостатков при построении комплексной системы верификации ключевых слов был использован классификатор на основе МОВ, обладающий значительными достоинствами при объединении мер достоверности.

Для верификации ключевых слов определим два класса. Если выбранный речевой фрагмент — ключевое слово, он относится к классу 1 $(y_1=+1)$, если фрагмент представляет ложную тревогу — то к классу 2 $(y_1=-1)$. В качестве входных параметров класса выберем меры доверительности CM_1, CM_2,… CM_u. При обучении системы одна из важных процедур — определение значение параметра c, а также выбор целевой функции f. В процессе обучения для ключевых и неключевых слов было предложено ввести различные весовые коэффициенты для c: параметры bC и $(1-b)C$ соответственно. В качестве целевой функции предложено использовать:

$f = \alpha k_1 - (1 - \alpha) k_2$, где k_1 — мера правильной классификации первого класса, k_2 — мера правильной классификации второго класса, $0,5 < \alpha < 1$ — пороговое значение.

Алгоритм верификации ключевых слов на основе мер достоверности с использованием МОВ дополнительно требует $O(mn^u)$ вычислений, где m — количество ключевых слов в словаре, n — размер выборки для обучения, u — количество мер достоверности. Поскольку число мер достоверности при верификации ключевых слов невелико, то вычислительная сложность алгоритма МОВ также достаточно невелика и лежит в пределах, позволяющих разрабатывать системы поиска в реальном масштабе времени. При правильном выборе функции ядра метод верификации ключевых слов на основе МОВ способен обеспечить высокую точность.

Гибридная система верификации ключевых слов

В результате декодирования на основе решётки и сети спутывания система ПКС генерирует оценки различных мер достоверности, сравнение которых с порогом позволяет принять решение о наличии или отсутствии искомых ключевых слов в потоке слитной речи. Однако использование такого подхода приводит к высокому уровню ложных тревог [8]. Уменьшение этого уровня за счёт изменения порога приводит к снижению вероятности правильного обнаружения.

Для устранения этого недостатка, присущего всем пороговым классификаторам, был использован гибридный подход на основе МОВ для верификации ключевых слов, позволяющий принимать многокритериальные решения. Этот подход реализован следующей структурой системы ПКС *(рис. 1)*.

Отличительная особенность предложенной структуры — то, что на выходе сети спутывания оцениваются меры достоверности и значений апостериорной вероятности, из анализа которых принимается решение о наличии пары спутывания. Если такая пара спутывания обнаруживается, то для вычисления апостериорной вероятности слогов спутывания используется гибридная модель [2].

Далее значения мер достоверности и апостериорных вероятностей поступают на вход блока многокритериального принятия решений, реализованного на основе МОВ.

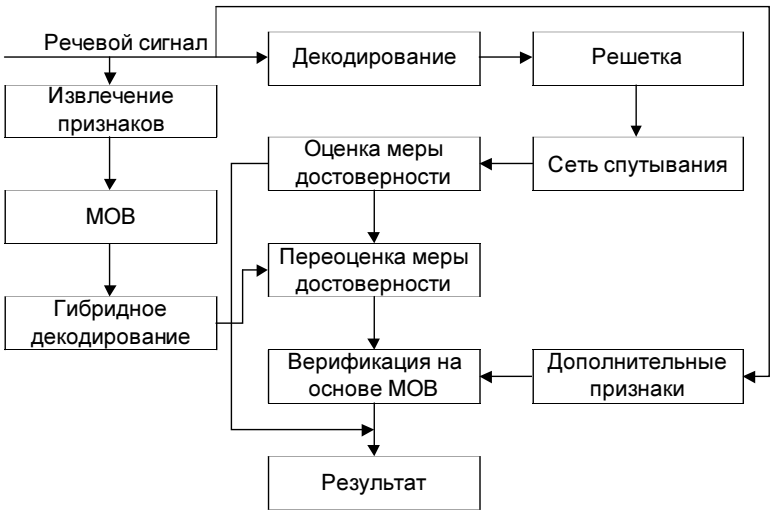

Рис. 1. Структура гибридной системы поиска ключевых слов

Алгоритм верификации ключевых слов на основе гибридной системы СММ-МОВ состоит из следующей последовательности действий:

1) Речевые данные поступают на систему декодирования, которая генерирует решётку слогов, которая затем преобразуется в сеть спутывания.

2) В каждом множестве спутывания фиксируются три слога с максимальными вероятностями и оценивается мера достоверности потенциального ключевого слова. Если полученная мера достоверности больше порогового значения, то происходит переход к этапу 4.

3) Создаётся объединённый вектор признаков, с помощью МОВ классификатора идентифицируются слоги спутывания, для которых переоценивается апостериорная вероятность.

4) Верификация ключевых слов.

Экспериментальное исследование

Для экспериментального исследования была составлена база речевых сигналов, содержащая записи русскоязычных интернет-радиостанций, общим объёмом 11.4 Гб. Для верификации ключевых слов использован классификатор на основе МОВ с ядром в виде гауссовой радиальной функции. Наилучшие характеристики верификации обеспечиваются при использовании следующих параметров классификатора на основе МОВ: $b = 0{,}96$, $\alpha = 0{,}6$ или $b = 0{,}99$, $\alpha = 0{,}6$. Для этих значений параметров проведено экспериментальное исследование по определению эффективности верификации на основе МОВ, объединяющей три различные меры доверительности — динамический рейтинг, нормализацию состояний и достоверность на интервале, так и для каждой меры доверительности в отдельности. Полученный результат приведён в *таблице 1*.

Таблица 1

Эффективность предложенного метода верификации ключевых слов

Pd,%	*FAR,%*			
	МОВ	Динамический рейтинг	Нормализация состояний	Достоверность на интервале
85.7	15.1	21.2	35.4	43.3

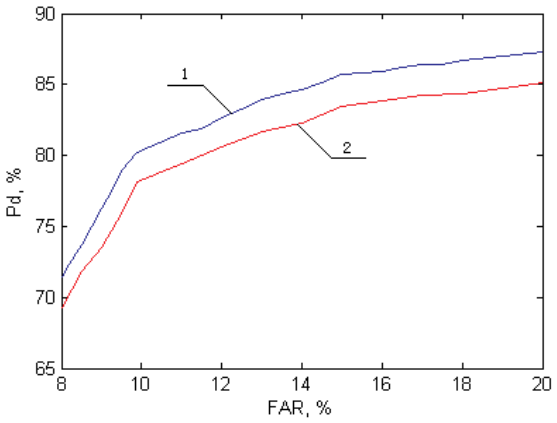

Рис. 2. Эффективность системы ПКС с использованием блока верификации ключевых слов (1) и без него (2)

Как видно из *таблицы 1*, для одних и тех же значений точности поиска Pd использование МОВ позволяет добиться меньших значений вероятности ложных тревог.

Разработанный на основе МОВ блок классификации использован для верификации ключевых слов в качестве дополнительной процедуры, на вход которой подаются результаты поиска ключевых слов на основе сети спутывания. На *рис. 2* представлена эффективность системы ПКС на основе сети спутывания, с дополнительным блоком верификации и без него.

Как видно из анализа рабочих характеристик, использование дополнительно блока верификации ключевых слов позволяет снизить вероятность ложной тревоги на 4.9% в зависимости от порога, используемого в сети спутывания.

Заключение

В статье рассмотрен метод верификации ключевых слов на основе метода опорных векторов, с использованием гибридной модели декодирования пары спутывания. Результаты экспериментального исследования показали, что использование блока верификации ключевых слов позволяет снизить уровень ложных тревог в среднем на 4.9% при фиксированной точности обнаружения.

Литература

1. *Рабинер Л.Р., Шафер Р.В.* Цифровая обработка речевых сигналов. М.: Радио и связь, 1981. С. 113–119.
2. *Сапожков М.А., Михайлов В.Г.* Вокодерная связь. М.: Радио и связь,1983. С. 156–158.
3. *Дегтярев Н.П.* Параметрическое и информационное описание речевых сигналов. Минск: Объединённый институт проблем информатики Национальной академии наук Беларуси, 2003. С.62–63.
4. *Лобанов Б.М., Цирульник Л.И.* Компьютерный синтез и клонирование речи. Минск: Белорусская наука, 2008. С.60-63.
5. *Быков Н.М. и др.* Надёжный метод выделения слоговых сегментов в речевом сигнале // Автоматика и информационно-измерительная техника. 2007. № 1.

Сорока Александр Михайлович —

научные интересы — методы и алгоритмы обработки цифровых сигналов, теория метода опорных векторов, смешанные гауссовы модели.