ИСПОЛЬЗОВАНИЕ ТРИГГЕРНОГО МОДЕЛИРОВАНИЯ ДЛЯ УЛУЧШЕНИЯ КАЧЕСТВА ПОИСКА КЛЮЧЕВЫХ СЛОВ В СЕТИ СПУТЫВАНИЯ

Янь Цзинбинь, Р.М. Алиев

Белорусский государственный университет, кафедра радиофизики
ул. Курчатова, 1, к. 52, г. Минск, Беларусь
телефон: + (37529) 5465716; e-mail: RomanAliyev@gmail.com

На данный момент не разработано эффективных алгоритмов поиска ключевых слов в сети спутывания, т. к. существуют проблемы, препятствующие этому. Первая проблема связана со сложностью создания сетей спутывания. Вторая проблема – влияние низких вероятностей появления гипотез в сети спутывания. В качестве решения первой проблемы предлагаются методы сегментации исходной решетки слов. Для решения второй проблемы используется переоценка лингвистической вероятности появления ключевого слова с помощью триггерных моделей.

Ключевые слова – поиск ключевых слов, сеть спутывания, триггерное моделирование.

1 ВВЕДЕНИЕ

Структура сети спутывания, получаемая путем извлечения информации из словесных структур, дает более ясное представление обо всех конкурирующих гипотезах и поможет улучшить точность распознавания речи. Такой подход дает возможность для анализа рядов конкурирующих гипотетических слов в предложении. Еще одно достоинство сети спутывания – это ее линейная структура, которую можно в любой момент дополнять новыми знаниями. В настоящее время решетка спутывания является самой компактной формой записи гипотетического результата распознавания непрерывной речи.

2 СЕТЬ СПУТЫВАНИЯ

Сеть спутывания представляет собой особый вид словесной структуры (рис. 1), которую можно рассматривать как направленный непериодический граф. Она задается, как $C=(Q,V,q_1,q_N,E)$, где N – количество узлов, E – множество всех связей, V – множество индексов всех связей, $Q=\{q_1,...,q_N\}$ - множество всех узлов, q_1 и q_N - начальный и конечный узлы сети. Каждая связь определяется четырьмя параметрами $e=(q_e^s,q_e^f,w_e,g_e)$, где $q_e^s,q_e^f\in Q$ - начальный и конечный узлы соединения e (если $q_e^s=q_i$, то $q_e^f=q_{i+1}$) w_e - индекс соединения, а $g_e\in(0,1)$ задает апостериорную вероятность возникновения соединения e. Узловое множество связей можно записать как $S_i^{cn}=\{e\,|\,e\in E,q_e^s=q_i\}\in E$, причем $\sum_{e\in S_i^{cn}}g_e=1$. Сеть спутывания, состоящая из N узлов, включает N-1 узловых множеств спутывания. Если *в узловом множестве существуют соединения с одинаковыми индексами*, то они объединяются в одно соединение.

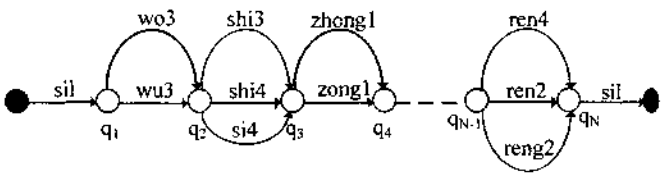

Рис. 1. Структура сети спутывания

Главная идея сети спутывания – обеспечить минимальный уровень словесных ошибок, что является самой важной целью при распознавании ключевых слов. Таким образом, применение решетки спутывания для распознавания ключевых слов имеет хорошее теоретическое обоснование и важное практическое значение.

3 ПРОЕКЦИОННЫЙ МЕТОД СОЗДАНИЯ СЕТИ СПУТЫВАНИЯ

Этот метод реализуется по следующему алгоритму:

1. Поиск проекций узлов решетки на временную ось (рис. 1);

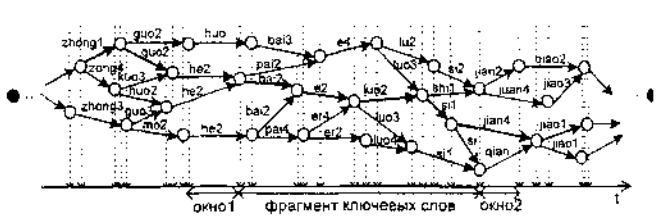

Рис. 1. Проекционный метод сегментирования решетки слов.

2. Анализ всех соединений в каждом временном окне, используя показатель сегмента G:

$$G = \frac{T_i}{T} D(E_i), \qquad (1)$$

где T_i - длина i-ого временного окна; T - общая длина речевого файла; E_i - множество соединений в i-ом временном окне. Временное окно задается отрезком между двумя соседними проекциями. Таким образом, чем выше среднее подобие индексов соединений и больше временная ширина окна, тем более вероятно получить решетку-сегмент, $D(E_i)$ – искажение выравнивания для E_i [1];

3. Исключая начальное и конечное окна, осуществляется выбор очередного временного окна с показателем сегмента, удовлетворяющему выражению:
$$G > G_{min}, \qquad (2)$$
G_{min} - пороговое значение показателя сегмента. Если такой выбор не возможен, то решетка не может быть дальше разделена и следует перейти к 5;
4. В текущем выбранном временном окне сконструировать решетку сегмент;
5. Если не конец, возвратится к 3, иначе завершить алгоритм.

4 ТРИГГЕРНОЕ МОДЕЛИРОВАНИЕ РЕЧИ

В любой произнесенной речи всегда можно выделить много контекстных взаимосвязей между словами. Известно, что человек воспринимает быстрее и лучше тесно связанную пару слов, чем слабо связанную. Контекстная взаимосвязь может наблюдаться между двумя существительными (например «день / ночь»), между прилагательным и существительным (например «громкий / голос»), в устойчивых выражениях (например «обращать / внимание») и др. Такие взаимосвязи можно получить из набора предварительно заданных данных и затем использовать для устранения неоднозначностей в сети спутывания.

На данный момент в распознавании речи преобладают простые n-грамные модели языка, которые могут учитывать близкую контекстную зависимость в пределах окна размером n слов (обычно n=3). Однако большинство зависимостей может оказаться за пределами этого окна. Таким образом, в n-грамных моделях фиксируются зависимости только на малых дистанциях.

Основываясь на вышесказанном, Розенфилд в качестве основной концепции для извлечения ассоциативной информации из связанной пары слов на малых и больших дистанциях использовал триггерную модель[2]. Структуру $(A_0 \rightarrow B)$ можно рассматривать как триггер, если инициируемое слово A_0 тесно связано с инициируемым словом B. Когда слово A_0 встречается в документе, оно инициирует слово B, вызывая переоценку вероятности его появления. Такой подход может быть расширен для длинных последовательностей слов.

Тогда моделирование речи проводится с помощью триггеров с использованием двух типов окна (Рис. 2).

Рис. 2. Расположение инициируемого ключевого слова в зависимости от типа окна: а – тип W1; б – тип W2.

Для окна типа W1 применяются следующие критерии для оценки вероятности инициализации ключевого слова w_k в триггере:

1. По максимальному значению вероятности (max(W1)):
$$\log P_T(w_k) = \log MAX[p(w_k \mid w_1), p(w_k \mid w_2), p(w_k \mid w_3), p(w_k \mid w_4)] + \log MAX[p(w_k \mid w_1 w_2), p(w_k \mid w_2 w_3), p(w_k \mid w_3 w_4)] \qquad (3)$$
2. По сумме вероятностей (sum(W1)):
$$\log P_T(w_k) = \log[p(w_k \mid w_1) + p(w_k \mid w_2) + p(w_k \mid w_3) + p(w_k \mid w_4)] + \log[p(w_k \mid w_1 w_2) + p(w_k \mid w_2 w_3) + p(w_k \mid w_3 w_4)] \qquad (4)$$
3. По среднему значению вероятностей (avg(W1)):
$$\log P_T(w_k) = \log\{[p(w_k \mid w_1) + p(w_k \mid w_2) + p(w_k \mid w_3) + p(w_k \mid w_4)]/4\} + \log\{[p(w_k \mid w_1 w_2) + p(w_k \mid w_2 w_3) + p(w_k \mid w_3 w_4)]/3\} \qquad (5)$$

Для окна типа W2 получаем:

1. По максимальному значению вероятности (max(W2)):
$$\log P_T(w_k) = \log MAX[p(w_k \mid w_1), p(w_k \mid w_2), p(w_k \mid w_3), p(w_k \mid w_4)] + \log MAX[p(w_k \mid w_2 w_3), p(w_k \mid w_4 w_5), p(w_k \mid w_5 w_6)] \qquad (6)$$
2. По сумме вероятностей (sum(W2)):
$$\log P_T(w_k) = \log[p(w_k \mid w_1) + p(w_k \mid w_2) + p(w_k \mid w_3) + p(w_k \mid w_6)] + \log[p(w_k \mid w_2 w_3) + p(w_k \mid w_4 w_5) + p(w_k \mid w_5 w_6)] \qquad (7)$$
3. По среднему значению вероятностей (avg(W2)):
$$\log P_T(w_k) = \log\{[p(w_k \mid w_1) + p(w_k \mid w_2) + p(w_k \mid w_3) + p(w_k \mid w_6)]/4\} + \log\{[p(w_k \mid w_2 w_3) + p(w_k \mid w_4 w_5) + p(w_k \mid w_5 w_6)]/3\} \qquad (8)$$

Для улучшения качества лингвистической модели, объединим биграмную и триггерную модель языка. При этом лингвистическая вероятность ключеого слова будет рассчитываться по следующей формуле:
$$P_l(w_k) = (1-\alpha) \cdot P_{2\text{-}g}(w_k \mid w_{k-1}) + \alpha P_T(w_k) \qquad (9)$$
где α<=1 – весовой коэффициент; $P_{2\text{-}g}$ – лингвистическая вероятность появления ключевого слова в биграмной модели.

5 ЭКСПЕРИМЕНТ

В ходе эксперимента были использованы инструментарии HTK[3] и SRLM[4].

В качестве экспериментальных данных было выбрано 100 предложений. Лингвистическая оценка ключевых слов задается формулой (9). На Рисунке 3 приведена зависимость значения вероятности обнаружения ключевых слов от значения весового коэффициента α. На данной зависимости ясно видно, что при использовании только триггерной модели языка (k=1) или только биграмной модели языка (k=0), получаемая вероятность не будет достигать максимального значения P. Оптимальное значение P получается при k=0,3. Этот эксперимент показал, что дополнительное использование триггерой модели языка приводит к небольшому увеличению точности, это обусловлено внесением в модель языка дополнительной контекстной информации о естественном языке.

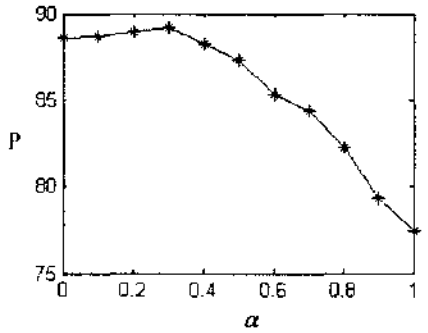

Рис. 3. Зависимость вероятности обнаружения ключевого слова от значения весового коэффициента α.

Так же было проведено экспериментальное сравнение различных видов триггерных моделей (табл. 1, 2):

ТАБЛИЦА 1

СРАВНЕНИЕ РАЗЛИЧНЫХ ВИДОВ ТРИГГЕРНЫХ МОДЕЛЕЙ ДЛЯ ОКНА ТИПА W1

	max(W1)	sum(W1)	avg(W1)
$P(w_k)$	89.1%	88.7%	88.8%
Уровень ошибок	21.1%	21.3%	21.1%

ТАБЛИЦА 2

СРАВНЕНИЕ РАЗЛИЧНЫХ ВИДОВ ТРИГГЕРНЫХ МОДЕЛЕЙ ДЛЯ ОКНА ТИПА W2

	max(W2)	sum(W2)	avg(W2)
$P(w_k)$	89.5%	89.5%	89.5%
Уровень ошибок	20.8%	21.0%	21.0%

Данный эксперимент показал, что эффективность поиска ключевых слов напрямую зависит от контекста, и второй вид триггерной модели (W2) имеет лучшие показатели.

6 ЗАКЛЮЧЕНИЕ

В этой статье был предложен новый метод создания сети спутывания на основе сегментации решетки слов проекционный метод, который позволяет уменьшить количество расчетов. А для повышения меры достоверности ключевых слов было предложено использовать триггерное моделирования языка. Результаты экспериментов показали улучшение вероятности обнаружения ключевых слов на 0,7 %.

ЛИТЕРАТУРА

[1] Wang, H.L. Research on Confusion Network and Side Information for Speech Recognition. /H.L Wang// Harbin Institute of Technology. Doctor's Degree. 2007. –P. 32-38

[2] Zhou GuoDong. Interpolation of n-gram and mutual-information based trigger pair language models for Mandarin speech recognition / GuoDong Zhou, KimTeng Lua // Computer Speech and Language. –1999. –vol.13. –P. 125–141.

[3] Young S. The HTK Book / S Young , G Evermann , D Kershaw //. http://htk.eng.cam.ac.uk.

[4] SRI language Modeling Toolkit. http://www.speech.sri.com/projects/srilm/.