

А. В. Максимов, С. Ю. Мельников, М. В. Федюкин, Н. М. Чавчавадзе (Москва, ТВП, «СТЭЛ КС»). О применении методов адаптации вероятностных моделей языка для повышения точности работы систем распознавания речи.

На рис. показана общая схема распознавателя речи, построенного с использованием аппарата скрытых марковских моделей [1]. Человек произносит некоторую фразу, которая представляет собой последовательность слов $W = w_1 w_2 \dots w_N$. Задача системы распознавания речи заключается в том, чтобы правильно распознать эту последовательность слов. Однако в ходе распознавания могут возникать ошибки, поэтому результат может оказаться отличным от W , например, $W' = w'_1 w'_2 \dots w'_M$. Для параметрического описания речевого сигнала он разделяется на короткие сегменты, для которых вычисляются вектора признаков $O = o_1 o_2 \dots o_T$.

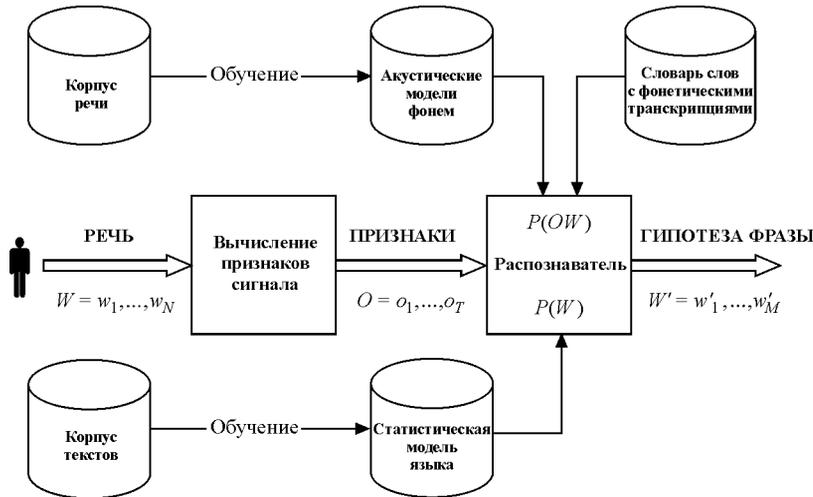


Рис.

Задачу распознавания сформулируем как задачу отыскания такой последовательности $W = w_1 w_2 \dots w_N$, для которой

$$P(w_1 w_2 \dots w_N | o_1 o_2 \dots o_T) P(w_1 w_2 \dots w_N) \rightarrow \max.$$

Здесь первый множитель отвечает за правдоподобие того, что при произнесении последовательности $O = o_1 o_2 \dots o_T$ получились наблюдения $W = w_1 w_2 \dots w_N$, а второй характеризует вероятность появления последовательности $W = w_1 w_2 \dots w_N$ в речи на данном языке.

Разработка систем дикторонезависимого распознавания речи требует наличия представительных речевых и текстовых корпусов для обучения лингвистических, т.е. акустических и статистических моделей языка. Объем корпусов, используемых для обучения моделей, напрямую влияет на точность создаваемых систем распознавания.

В некоторых случаях при работе с редкими языками разработчик при создании акустических моделей ограничен имеющейся в его распоряжении речевой базой, но имеет возможность увеличивать объем текстовой базы для построения более точных вероятностных лингвистических моделей. Более того, в таких приложениях, как мониторинг аудиовизуальных СМИ (в англоязычной литературе «Broadcast Recognition») может быть доступной информация о тематике речевого сообщения, например, когда имеются варианты заголовка сообщения в текстовом виде.

В докладе обсуждается возможность использования методов адаптации лингвистических моделей в описанной ситуации для повышения точности распознавания.

Основная идея предлагаемого подхода — по имеющемуся варианту заголовка осуществить подбор текстов по данной тематике, провести адаптацию статистической модели языка по данной тематике, адаптировать акустическую модель и осуществить процесс распознавания с применением модернизированной лингвистической модели. При этом подбор текстов может осуществляться как из имеющейся базы, так и из новостных лент с использованием web-ресурсов.

Процесс адаптации статистических моделей может осуществляться несколькими способами, описанными в литературе [2–4]. В случае, когда используются статистические модели на n -граммах слов, для повышения скорости вычислений можно использовать адаптацию на основе линейной интерполяции: $P_{\text{адаптир}}(w, \text{history}) = \sum_i \lambda_i P_i(w, \text{history})$, где λ_i — веса используемых моделей, w — слово, вероятность которого оценивается, history — набор предшествующих слов.

Процесс адаптации акустических моделей заключается в переоценке вероятностей таких фонетических единиц речи, как бифоны, трифоны и другие фонетические конструкции.

СПИСОК ЛИТЕРАТУРЫ

1. *Rabiner L., Juang B.* Fundamentals of Speech Recognition. New Jersey: Prentice-Hall, Englewood Cliffs, 1993.
2. *Bacchiani M., Roark B.* Unsupervised language model adaptation. — In: Proceedings of the International Conference on Acoustics, Speech, and Signal Processing, 2003, p. 224–227.
3. *Sepesy Maucec M., Kacic Z.* Topic detection for language model adaptation of highly-inflected languages by using a fuzzy comparison function. — In: Proceedings of 7th European Conference on Speech Communication and Technology, 2001, v. A42, p. 243–247.
4. *Martin S. C., Liermann J., Ney H.* Adaptive topic-dependent language modelling using word-based varigrams. — In: Proceedings of 5th European Conference on Speech Communication and Technology, 1997, p. 1447–1450.