

М.Ю. Зулкарнеев, С.Х. Сальман, Н.Г. Шамраев. Ростов-на-Дону, ФГНУ НИИ «Спецвузавтоматика». **Использование метода полусвязанных ковариационных матриц для распознавания речи.**

В настоящее время наилучшие результаты в области автоматического распознавания речи показывают системы, использующие скрытые марковские модели (СММ), со взвешенными смесями нормальных (гауссовых) распределений в качестве выходных вероятностей [1]. Обычно в таких системах в качестве ковариационных матриц используются диагональные. При использовании полных матриц ковариаций, значительно увеличивается число параметров, моделирующих один компонент, при этом ограничивается количество компонент, которые можно надежно оценить.

Для решения этой проблемы в настоящей работе используется метод полусвязанных ковариационных матриц [2]. Рассмотрим компоненты с полными ковариационными матрицами  $\Sigma^{(m)}$ . Для каждого компонента существует пространство, в котором его матрица является диагональной  $\Sigma_{\text{diag}}^{(m)}$ . Пусть  $H^{(m)}$  преобразование между этими пространствами:

$$\Sigma^{(m)} = H^{(m)} \Sigma_{\text{diag}}^{(m)} H^{(m)}.$$

В соответствии с работой [2] разобьем все компоненты на классы и для каждого класса используем свою матрицу преобразований  $H^{(r)}$ ,  $r = 1 \dots R$ ,  $R$  — количество классов. В этом случае для  $M$  компонент с полными ковариационными матрицами потребуется  $dM + \frac{d}{2}(d+1)R$  параметров вместо  $\frac{d}{2}(d+1)M$ . Это может значительно понизить количество оцениваемых параметров, и при этом учитывать ковариацию при моделировании отдельных компонент смесей.

Для оценки параметров компонент и матриц  $H^{(r)}$  применяется ЕМ (expectation-maximisation) подход [2] со вспомогательной функцией

$$Q(M, \bar{M}) = \sum_{m \in M^{(r)}, \tau} \gamma_m(\tau) \left( \frac{|\bar{A}^{(r)}|^2}{|\bar{\Sigma}_{\text{diag}}^{(m)}|} \right) - (o(\tau) - \bar{\mu}^{(m)})^T \bar{A}^{(r)T} \bar{\Sigma}_{\text{diag}}^{(m)-1} \bar{A}^{(r)} (o(\tau) - \bar{\mu}^{(m)}),$$

где  $|\cdot|$  — означает определитель матрицы,  $M^{(m)}$  — множество гауссовых компонент, относящихся к классу  $r$ .

Используя полусвязанные ковариационные матрицы, были проведены эксперименты по распознаванию ключевых слов в слитной русской речи. Точности распознавания ключевых слов (отношение количества правильно найденных слов к количеству ключевых слов в тестирующем материале) приведены в табл. Аналогичный эксперимент без применения полусвязанных ковариационных матриц позволил получить точность 89,5%. Применение метода полусвязанных ковариационных матриц позволило повысить точность распознавания ключевых слов на 2,2%.

**Таблица.** Точность распознавания в зависимости от количества классов

Количество классов	2	4	8	16	32	64	128
Точность, %	90,01	90,71	91,56	91,35	91,56	91,77	91,77

#### СПИСОК ЛИТЕРАТУРЫ

1. *Rabiner L. R.* A tutorial on hidden Markov models and selected applications in speech recognition applications. — Proceedings of the IEEE, 77, February, 1989.
2. *Gales M. J.* Semi-tied covariance matrices for hidden Markov models. Cambridge University, February, 1998.
3. *Demster A. P., Laird N. M., Rubin D. B.* Maximum likelihood from incomplete data via the EM algorithm. — Journal of the Royal Statistical Society, 39:1-38, 1977.