

ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДОРОЖНОГО ДВИЖЕНИЯ: СОСТОЯНИЕ И ПЕРСПЕКТИВЫ РАЗВИТИЯ

УДК 004

РАЗРАБОТКА ДЕТЕРМИНИРОВАННОГО СТРУКТУРНОГО МЕТОДА РАСПОЗНАВАНИЯ СЛОВ КОМАНД ВОДИТЕЛЯ

Е.Е. Федоров, доцент, к.т.н., Донецкая академия автомобильного транспорта

***Аннотация.** Предлагается детерминированный структурный метод распознавания слов команд водителя, который использует признаки, полученные путем нерасширяющих равномерно непрерывных отображений, и детерминированный конечный автомат, который моделирует слова команд водителя.*

***Ключевые слова:** детерминированный конечный автомат, распознавание слов команд, компьютерная система, нерасширяющие равномерно непрерывные отображения, мел-частотные кепстральные коэффициенты.*

РОЗРОБКА ДЕТЕРМІНОВАНОГО СТРУКТУРНОГО МЕТОДУ РОЗПІЗНАВАННЯ СЛІВ КОМАНД ВОДІЯ

Є.Є. Федоров, доцент, к.т.н., Донецька академія автомобільного транспорту

***Анотація.** Пропонується детермінований структурний метод розпізнавання слів команд водія, що використовує ознаки, отримані шляхом нерозширювальних рівномірно безперервних відображень, і детермінований кінцевий автомат, який моделює слова команд водія.*

***Ключові слова:** детермінований кінцевий автомат, розпізнавання слів команд, комп'ютерна система, нерозширювальні рівномірно безперервні відображення, мел-частотні кепстральні коефіцієнти.*

DEVELOPING OF DETERMINED STRUCTURAL METHOD OF DRIVER'S COMMAND WORDS RECOGNITION

**E. Fiodorov, Associate Professor, Candidate of Technical Science,
Donetsk Academy of Motor Transport**

***Abstract.** The structural method of command words recognition of the driver which uses the features received by uniformly expanding continuous mapping, and the determined finite state machine, which models words of commands of the driver is offered.*

***Key words:** determined finite state machine, commands words recognition, computer system, uniformly expanding continuous mapping, mel-frequency cepstral coefficients.*

Введение

В настоящее время широкое распространение получают системы человеко-машинного общения, в частности, для управления транспортными объектами. Важную роль в таких системах играет распознавание речевых ко-

манд. Для эффективного распознавания вербальных команд водителя принятие решений должно производиться с высоким быстродействием и высокой вероятностью. Параметры системы идентификации команд водителя определяются по результатам численного исследования.

Анализ публикаций

Существующие методы и модели распознавания речевых образов обычно основаны на скрытых марковских моделях (СММ) [1], алгоритме динамического программирования DTW [2] и нейронных сетях [3–6] и обладают одним или несколькими из перечисленных ниже недостатков:

- длительность обучения;
- хранение большого количества эталонов звуков или слов;
- хранение большого количества весовых коэффициентов;
- длительность распознавания;
- низкая вероятность распознавания;
- необходимость большого количества обучающих данных.

Цель и постановка задачи

Цель статьи – для вербального управления транспортными объектами разработать детерминированный структурный метод распознавания слов команд водителя, который предусматривает решение следующих задач:

- формирование признаков посредством нерасширяющих равномерно непрерывных отображений;
- построение детерминированного конечного автомата, описывающего слова команд.

Формирование признаков посредством нерасширяющих равномерно непрерывных отображений

В работе [7] впервые были введены нерасширяющие равномерно непрерывные отображения, действующие в компактных метрических пространствах образцов сигналов. В данной статье эти отображения используются с точки зрения преобразования и распознавания образцов звуков речи и исследуются на мел-частотных кепстральных коэффициентах (MFCC), которые описаны в работе [8].

Нерасширяющее равномерно непрерывное отображение $\phi: S^N \rightarrow V^K$ соответствует функции формирования вектора признаков, т.е. отображает вектор целых значений дискретного сигнала $\mathbf{s} = (s_1, \dots, s_N)$ в вектор вещественных признаков $\mathbf{v} = (v_1, \dots, v_K)$, причем каждая компонента вектора \mathbf{v} вычисляется как MFCC в виде

$$v_i = \phi_i(\mathbf{s}) = MFCC_{i-1}, i \in \overline{1, K},$$

$$MFCC_0 = \frac{1}{P} \sum_{i=1}^P E_i,$$

$$MFCC_j = \sum_{i=1}^P E_i \cos(j(i - 0.5)\pi / P), j \in \overline{1, K - 1}$$

$$E_i = \lg \left(\sum_{k=k1_i}^{k2_i} (X(k))^2 w2(k) \right), i \in \overline{1, P}$$

$$X(k) = \sum_{m=0}^{N-1} (s_{m+1} - 2^{r_s}) w1(m) e^{-j(2\pi / N)km}$$

$$w1(m) = 0,54 + 0,46 \cos \frac{2\pi m}{N},$$

$$w2(k) = \begin{cases} 0, & k < k1_i \\ \frac{k - k1_i}{\Delta k_i / 2}, & k1_i \leq k \leq k1_i + \Delta k_i / 2 \\ \frac{k2_i - k}{\Delta k_i / 2}, & k1_i + \Delta k_i / 2 \leq k \leq k2_i \\ 0, & k > k2_i \end{cases},$$

где E_i – логарифмированная энергия i -й мел-частотной полосы Фурье-спектра; $X(k)$ – Фурье-спектр сигнала; $w1(m)$ – окно Хемминга; $w2(k)$ – треугольное окно; P – количество мел-частотных полос спектра; r_s – количество разрядов (бит) для одного значения сигнала; $k1_i, k2_i$ – границы частотных диапазонов i -й полосы, $\Delta k_i = k2_i - k1_i$.

Нерасширяющее равномерно непрерывное отображение $\psi: V^K \rightarrow X^K$ соответствует функции нормирования, масштабирования и округления вектора вещественных признаков, т.е. отображает вектор вещественных признаков $\mathbf{v} = (v_1, \dots, v_K)$ в вектор целых признаков $\mathbf{x} = (x_1, \dots, x_K)$, причем каждая i -я компонента вектора \mathbf{x} вычисляется в виде

$$x_i = \psi_i(\mathbf{v}) = \left[\frac{v_i - \min_{k \in \overline{1, K}} v_k}{\max_{k \in \overline{1, K}} v_k - \min_{k \in \overline{1, K}} v_k} \cdot \alpha \right], i \in \overline{1, K},$$

где $\alpha \in \{2^{r_x-1}, \dots, 2^{r_x}\}$ – коэффициент масштаба, $\lceil \cdot \rceil$ означает округление; r_x – количество разрядов (бит) для одного значения признака.

Неразширяющее равномерно непрерывное отображение $\varphi: X^K \rightarrow Y^M$ соответствует функции классификации вектора целых признаков, т.е. отображает вектор целых признаков $\mathbf{x} = (x_1, \dots, x_K)$ в номер класса образца звука речи, представленный булевым вектором с одной ненулевой компонентой $\mathbf{y} = (y_1, \dots, y_M)$, причем каждая i -я компонента вектора \mathbf{y} вычисляется в виде

$$y_i = \varphi_i(\mathbf{x}) = \begin{cases} 1, & \inf_{\tilde{\mathbf{x}} \in X_i^K} \rho(\mathbf{x}, \tilde{\mathbf{x}}) = 0 \\ 0, & \inf_{\tilde{\mathbf{x}} \in X_i^K} \rho(\mathbf{x}, \tilde{\mathbf{x}}) = 1 \end{cases}, i \in \overline{1, M},$$

$$\rho(\mathbf{x}, \tilde{\mathbf{x}}) = \begin{cases} 1, & d(\mathbf{x}, \tilde{\mathbf{x}}) \geq 0 \\ 0, & d(\mathbf{x}, \tilde{\mathbf{x}}) < 0 \end{cases},$$

$$d(\mathbf{x}, \tilde{\mathbf{x}}) = \sum_{k=1}^K |x_k - \tilde{x}_k|,$$

где X_i^K – класс эквивалентности, соответствующий i -му звуку речи.

На рис.1 приведены обучающие образцы фонемы слова «на» после выполнения нормирования, масштабирования и округления. Каждый образец представляет собой вектор признаков, вычисленный на основе мел-частотных кепстральных коэффициентов (MFCC).

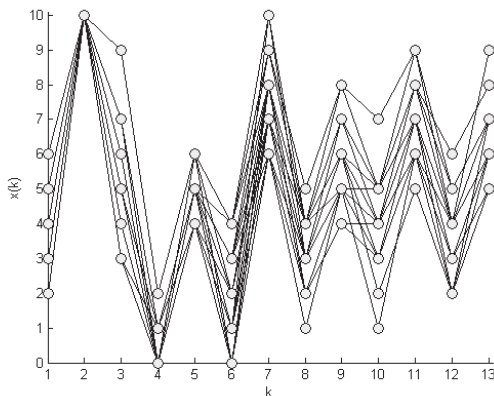


Рис. 1. Обучающие образцы фонемы

Построение детерминированного конечного автомата, описывающего слова команд

Пусть с одной стороны распознаваемому слову команды соответствует конечная последовательность фонем \mathbf{W} длины $|w|$, которая состоит из M классов различных

фонем, $|w| \geq M$. С другой стороны распознаваемому слову соответствует цифровой сигнал, который разбит на T перекрывающихся фреймов длиной N , причем каждый последующий фрейм сдвинут относительно предыдущего фрейма на один отсчет. Произведем маркировку фреймов цифрового сигнала следующим образом. Пусть t -му фрейму слова, $1 \leq t < T$, соответствует символ c_0 , если в результате распознавания этот фрейм не был классифицирован (т.е. образец t -го фрейма не совпал ни с одним из опорных образцов фонем); t -му фрейму слова, $1 \leq t < T$, соответствует символ c_k , $k \in \overline{1, M}$, если в результате распознавания этот фрейм был отнесен к k -му классу (т.е. образец t -го фрейма совпал с одним из опорных образцов фонем); T -му фрейму слова соответствует символ c_{M+1} . В результате будет получена конечная последовательность маркеров фреймов $\mathbf{Q} = q_1, q_2, \dots, q_T$, где

$q_t \in \{c_0, c_1, \dots, c_M\}$, $t \in \overline{1, T-1}$ и $q_T = c_{M+1}$.

Чтобы полученная последовательность \mathbf{Q} была в дальнейшем правильно преобразована к последовательности \mathbf{W} , необходимо наложить на последовательность \mathbf{Q} ограничения посредством следующего утверждения.

Утверждение. В конечной последовательности маркеров фреймов $\mathbf{Q} = q_1, q_2, \dots, q_T$ распознаваемого слова символ c_k , соответствующий фонеме k -го класса этого слова, должен подчиняться следующим условиям:

1. Необходимое условие. Символ c_k не должен появляться для фонем не k -го класса распознаваемого слова.
2. Достаточное условие. Символ c_k должен появляться хотя бы один раз для каждой фонемы k -го класса распознаваемого слова.

Необходимое условие налагает ограничение на типы преобразований фреймов сигнала, используемых при формировании опорных образцов фонем. Достаточное условие налагает ограничение на количество опорных образцов фонем.

Сформулированное утверждение можно представить в виде следующего логико-формального правила

$$\forall i \in \overline{1, |w|}, j \in \overline{1, |w|} \exists k :$$

$$fonem_i \in \mathbf{W} \wedge fonem_j \in \mathbf{W} \wedge$$

$$\wedge class(fonem_i) = k \wedge lbound(fonem_i) = n \wedge$$

$$rbound(fonem_i) = m \wedge class(fonem_j) \neq k \wedge$$

$$\wedge lbound(fonem_j) = s \wedge rbound(fonem_j) = r \rightarrow$$

$$\rightarrow \left(\bigvee_{l=n}^m (q_l = c_k \wedge q_l \in \mathbf{Q}) \right) \wedge$$

$$\wedge \left(\bigwedge_{p=s}^r (q_p \neq c_k \wedge q_p \in \mathbf{Q}) \right),$$

где *class* – функция, возвращающая номер класса фонемы;

lbound – функция, возвращающая номер фрейма, содержащего начало фонемы;

rbound – функция, возвращающая номер фрейма, содержащего конец фонемы.

Для генерации последовательности **Q** можно предложить следующую генеративную модель в виде детерминированного конечного автомата:

$$DFA = (U, C, \delta, u_0, U_{end}),$$

где *U* – конечное множество состояний; $U = \{u_0, u_1, \dots, u_{2M+2}\}$; *C* – конечное множество допустимых входных символов; $C = \{c_0, c_1, \dots, c_{M+1}\}$; δ – функция переходов, $\delta: U \times C \rightarrow U$; т.е. $\delta(u_i, c_k) = u_j$, u_0 – начальное состояние; U_{end} – множество заключительных состояний, $U_{end} = \{u_{2M+2}\}$.

Пример детерминированного конечного автомата для слова «на», где $M = 2$, представлен в виде графа на рис. 2.

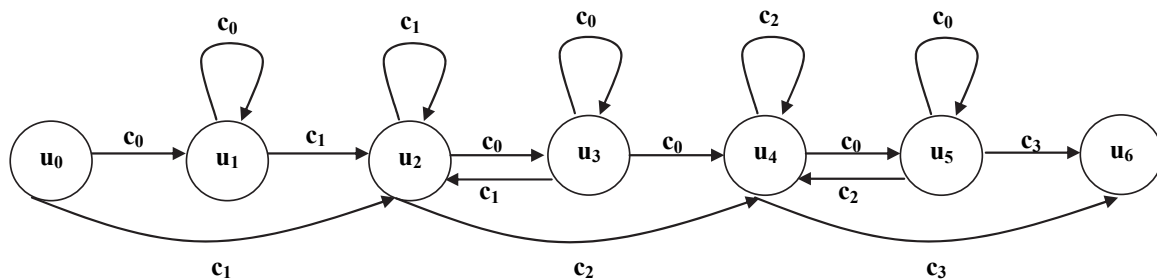


Рис. 2. Детерминированный конечный автомат в виде графа для слова «на»

Выводы

Предложенный детерминированный структурный метод обладает следующими достоинствами:

1. Допускается, что некоторые фреймы могут быть не классифицированы. Это прежде всего относится к фреймам, принадлежащим двум соседним фонемам.
2. Сокращается время формирования опорных образцов для фонемы и уменьшается объем хранимой информации, поскольку становится возможным ограничиться только частью опорных образцов данной фонемы.
3. Не требуется вычислять вероятности переходов между состояниями (как, например, для скрытых марковских моделей), а также

вычислять минимальное и максимальное количество повторений фреймов в слове (как, например, для подхода КДП).

4. Модель позволяет распознавать быстро произносимые слова, фонемы которых имеют малую длину.

Предложенный метод может использоваться в интеллектуальных компьютерных системах.

Литература

1. Rabiner L.R. Fundamentals of speech recognition / L.R. Rabiner, B.H. Jang. – Englewood Cliffs, NJ: Prentice Hall PTR, 1993. – 507 p.
2. Винцок Т.К. Анализ, распознавание и интерпретация речевых сигналов / Т.К. Винцок. – К.: Наук. думка, 1987. – 260 с.

3. Осовский С. Нейронные сети для обработки информации / С. Осовский. – М.: Финансы и статистика, 2002. – 344 с.
4. Хайкин С. Нейронные сети: полный курс / С. Хайкин. – М.: Издательский дом «Вильямс», 2006. – 1104 с.
5. Комарцова Л.Г. Нейрокомпьютеры / Л.Г. Комарцова, А.В. Максимов. – М.: Изд-во МГТУ имени Н.Э. Батмана, 2002. – 320 с.
6. Каллан Р. Основные концепции нейронных сетей / Р. Каллан. – М.: Издательский дом «Вильямс», 2001. – 288 с.
7. Федоров Е.Е. Метод обработки сигнала на основе нерасширяющих равномерно непрерывных отображений / Е.Е. Федоров // Радіотехнічні поля, сигнали, апарати та системи (теорія, практика, історія, освіта): труды Междунар. науч.-техн. конф. (РТПСАС'2012). – К., 2012. – С. 343–346.
8. Davis S. Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Sentences / S. Davis, P. Mermelstein // IEEE Trans. on Acoustic, Speech and Signal Processing. – 1980. – Vol.28, №4. – P. 357–366.

Рецензент: Л.С. Абрамова, доцент, к.т.н., ХНАДУ.

Статья поступила в редакцию 11 марта 2013 г.
