ИНФОРМАТИКА И СИСТЕМЫ УПРАВЛЕНИЯ

УДК 681.391

М. О. Дербасов[2], А. С. Лаптев[3], А. А. Филяков[1], В. Е. Гай[1]

РЕЧЕВОЕ УПРАВЛЕНИЕ РОБОТОТЕХНИЧЕСКОЙ СИСТЕМОЙ С ПОЗИЦИИ ТЕОРИИ АКТИВНОГО ВОСПРИЯТИЯ

Нижегородский государственный технический университет им. Р. Е. Алексеева[1],
ЗАО «Интел»[2],
Нижегородский радиотехнических колледж[3]

Работа посвящена описанию метода распознавания речевых команд в условиях априорной неопределенности в задачах управления робототехнической системой с позиции активного восприятия. В отличие от существующих методов распознавания, работающих на уровне отсчётов, предлагаемый метод реализует концепцию грубо-точного анализа сигнала, описанную в теории активного восприятия.

Ключевые слова: распознавание голосовых команд, теория активного восприятия.

Введение

Робототехническая система является сложным программно-аппаратным комплексом, активно взаимодействующим с внешней средой. Структурно это взаимодействие можно представить в виде схемы, представленной на рис. 1.

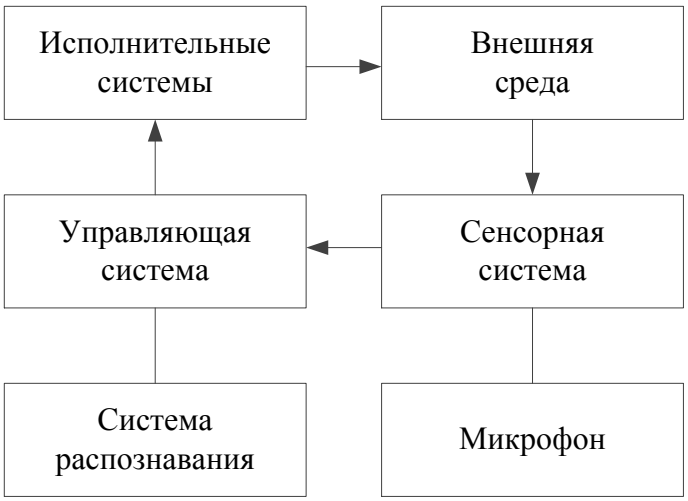

Рис. 1. Структурное представление взаимодействий

Задача ручного управления робототехнической системой еще полностью не решена. Основной проблемой является отсутствие удобного и простого в обращение, особенно для неподготовленного оператора, пользовательского интерфейса. Даже для проведения простых манипуляций с физическими предметами может потребоваться несколько десятков комбина-

ций команд. Создание же интерфейса, понятного и эргономичного для человека, в таком ключе становиться почти невыполнимой задачей.

Одним из решений данной проблемы может стать управление устройством при помощи речевых команд, подаваемых человеком. Существует несколько больших классов методов распознования речи: скрытые Марковские модели; нейронные сети; методы дискриминантного анализа, основанные на Байесовской дискриминации; динамическое программирование – временные динамические алгоритмы, каждый из которых имеет свои достоинства и недостатки. В данной статье будет рассматриваться процесс распознования с позиции системного анализа.

Процесс распознавания с позиций системного анализа можно разделить на три этапа: формирование исходного описания, нахождение системы признаков и построение решающего правила. Существуют две формулировки задачи распознавания: в узком и широком смыслах [1]. В узком смысле задача распознавания сводится к построению классификатора, в широком – к распознаванию в условиях априорной неопределённости (в данном случае не известны множество признаков и множество классов).

Известны проблемы, связанные с применением существующих методов распознавания образов [2]:

1) проблема формирования исходного описания. Связана с тем, что существующие модели и методы распознавания адаптированы к конкретному классу прикладных задач и требуют априорного знания свойств анализируемых сигналов;

2) проблема формирования системы признаков. Связана с выбором конечного множества признаков, обеспечивающих однозначность решения задачи классификации на этапе распознавания и отвечающая требованиям необходимости и достаточности. Этап выбора системы признаков необходим для сокращения размерности входного описания. Поскольку задача сокращения размерности – оптимизационная задача, то для её решения следует использовать критерий информативности. Отсутствие модели априорной неопределённости и модели её раскрытия породило большое количество методов в выборе критерия информативности, что привело к большому числу возможных вариантов признаков [3, 4];

3) проблема принятия решений в условиях априорной неопределённости. Этап принятия решения заключается в сравнении с имеющимся эталоном признакового описания анализируемого сигнала. Предполагается, что эталону соответствует компактное множество точек в системе признаков. Однако помехи, структурные изменения одного и того же представителя класса приводят к перекрытию классов. Поэтому проблема принятия решения замыкается на проблеме формирования системы признаков, позволяющей сформировать эталон, имеющий компактное представление.

Теория активного восприятия предлагает решение описанных проблем [1]. Настоящая работа посвящена применению данной теории к анализу речевых сигналов для управления робототехнической системой.

1. Обзор методов распознавания речевых сигналов

Рассмотрим методы, применяемые на разных этапах решения задачи распознавания [5]:

1) этап предварительной обработки звукового сигнала. Обычно он заключается в фильтрации сигнала и выделении границ речевой активности [6, 7]. Учитывая, что задача распознавания решается в условиях априорной неопределённости (информация о помехе отсутствует), выбрать подходящий фильтр сложно;

2) для создания описания входного сигнала вычисляются признаки: коэффициенты спектра Фурье; кепстральные коэффициенты; мел-частотные кепстральные коэффициенты; коэффициенты линейного предсказания (linear predictive coding); коэффициенты вейвлет-спектра и т. д. Необходимо отметить, что существующие методы обработки речевых сигналов основаны на стратегии точно-грубого анализа, который заключается в том, что признаки вычисляются по участку сигнала длительностью около 25 мс [4, 5];

3) на этапе классификации в системах распознавания речи взаимодействуют несколько модулей [8]:

а) модуль акустической модели позволяет по входному речевому сегменту определить наиболее соответствующие ему шаблоны отдельных звуков. При акустическом моделировании используются скрытая марковская модель, модель гауссовой смеси, нейронная сеть, метод опорных векторов. Применение данных моделей предполагает их предварительное обучение и выбор параметров, что, в условиях априорной неопределённости является не тривиальной работой;

б) модуль модели языка служит для определения наиболее вероятной последовательности слов. Необходимость использования языковой модели объясняется ростом словаря распознаваемых слов, в результате чего увеличивается число слов, похожих по звучанию. Выделяют дискретные (модель с конечным числом состояний, на основе теории формальных языков, на основе лингвистических знаний) и статистические модели (*n*-граммная модель, модель на основе деревьев решений, статистическое обобщение формальных языков);

в) декодер объединяет данные, поступающие от акустической и языковой моделей, и формирует результат распознавания.

2. Метод распознования речевых сигналов на основе теории активного восприятия

В теории активного восприятия (ТАВ) описан метод грубо-точного анализа, используемый для распознавания изображений. Предполагается, что похожие механизмы работают в слуховой системе, исходя из чего данный метод может быть применён и к анализу речевых сигналов. Рассмотрим предлагаемую реализацию этапов системы распознавания с точки зрения ТАВ.

2.1. Предварительная обработка

В условиях априорной неопределённости процесс раскрытия неопределённости звукового сигнала заключается в дихотомии его области определения G на равные части. Поскольку все отсчёты сигнала находятся в отношении эквивалентности, множество отсчётов можно разбить на любое число подобластей $G_{ij} \subseteq G$ без пересечения этих областей между собой. Последовательное применение операции дихотомии позволяет сгенерировать пирамидальную структуру (рис. 2).

Таким образом, этап предварительной обработки заключается в выполнении операции дихотомии и формировании подобластей G_{ij}.

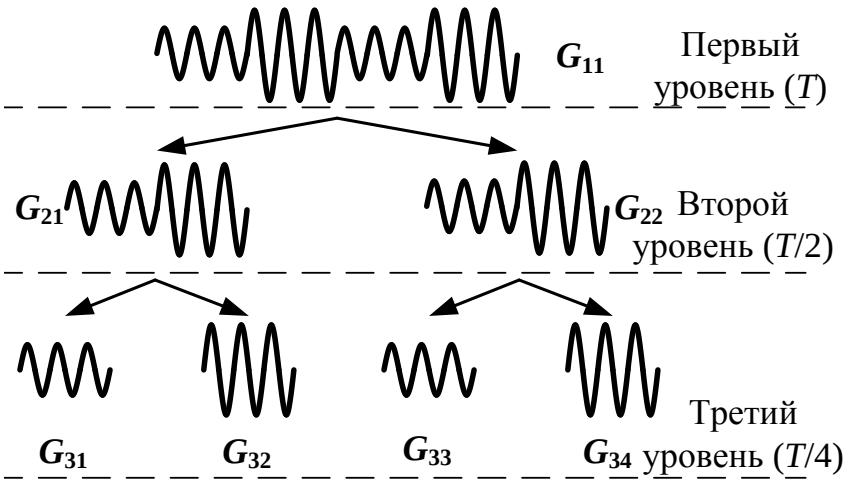

Рис. 2. Пирамида описания сигнала:
i – уровень разложения; *j* – номер области на *i*-м уровне;
T – длительность сигнала

2.2. Вычисление признаков

Рассмотрим предлагаемый метод вычисления признакового описания подобласти $G_{ij} \subseteq G$:

1) отсчёты сигнала, относящиеся к подобласти G_{ij}, разбиваются на множество сегментов $\mathbf{g} = \{g_k\}$ длиной $L * 16$ отсчётов со смещением в S отсчётов, $k = \overline{1, N}$, где N – число сегментов в подобласти G_{ij};

2) к каждому сегменту g_k применяется U-преобразование (U-преобразование является базовым в теории активного восприятия), в результате формируется спектральное представление каждого сегмента $u_k = U[g_k]$, $\boldsymbol{u} = \{u_k\}$, где U – оператор вычисления U-преобразования;

3) по вычисленному спектральному представлению u_k сегмента g_k определяются замкнутые группы $p_k = P[u_k]$, $\boldsymbol{p} = \{p_k\}$, где P – оператор вычисления замкнутых групп;

4) вычисляется гистограмма замкнутых групп $d_{ij} = H[\boldsymbol{p}]$, где H – оператор формирования гистограммы замкнутых групп, которая и является признаковым описанием области G_{ij};

5) признаковые описания областей G_{ij} объединяются в вектор x.

Отметим, что при создании признакового описания используется принцип рекурсии, т. е. к сигналу последовательно применяется одна и та же операция – дихотомия. Таким образом, для выявления структуры сложного сигнала применяется одна и та же операция.

2.3. Принятие решения (классификация)

Этап классификации может быть реализован с помощью нескольких классификаторов. В данной работе используется линейный метод опорных векторов (SVM), также известный под названием метод классификации с максимальным зазором. Основная идея этого метода заключается в переводе исходных векторов в просторанство более высокой размерности и поиск разделяющих гиперплоскости с максимальным зазором в этом пространстве. Две параллельные гиперплоскости строятся по обе стороны от гиперплоскости, разделяющей конечные классы. Разделяющей гиперплоскостью будет гиперплоскость, максимизирующая расстояние до двух параллельных гиперплоскостей. Метод работает в предположение, что чем больше разница или расстояние между этими параллельными гиперплоскостями, тем меньше будет средняя ошибка классификатора.

Входные данные для классификатора могут пройти предварительную нормализацию, но это требует дополнительных вычислительных ресурсов.

Решающее правило метода опорных векторов выглядит следующим образом:

$$a(x) = \text{sign}\left(\sum_{j=1}^{n} w_j x^j - w_0\right),$$

где $x = \left(x^1, \ldots, x^n\right)$ – признаковое описание объекта x (одно из возможных описаний, приведённых выше); вектор $w = \left(w^1, \ldots, w^n\right)$ и скалярный порог w_0 являются параметрами алгоритма. Метод опорных векторов является бинарным классификатором. В данной работе для решения задачи мультиклассовой классификации используются два способа сведения данной задачи к бинарной [5]:

1) подход «один против всех» (One-vs-All) заключается в обучении N классификаторов по следующему принципу:

$$f_i(x) = \begin{cases} \geq 0, \text{если } y(x) = i, \\ < 0, \text{если } y(x) \neq i, \end{cases}$$

вычисляются все классификаторы и выбирается класс, соответствующий классификатору с большим значением

$$a(x) = \arg\max_{i \in \overline{1,N}} f_i(x);$$

2) подход «один против одного» (One-vs-One) заключается в формировании $N(N-1)$ классификаторов, которые разделяют объекты пар различных классов,

$$f_{ij}(x) = \begin{cases} +1, если\ y(x) = i, \\ -1, если\ y(x) = j. \end{cases}$$

После обучения бинарных классификаторов, решение принимается следующим образом:

$$a(x) = \arg\max_{i \in \overline{1,N}} \sum_{\substack{j=1 \\ j \neq i}}^{N} f_{ij}(x).$$

При классификации используется линейное ядро $k(x,y) = x^T y + c.$

Структурная схема системы класификации принятого сигнала представлена на рис. 3.

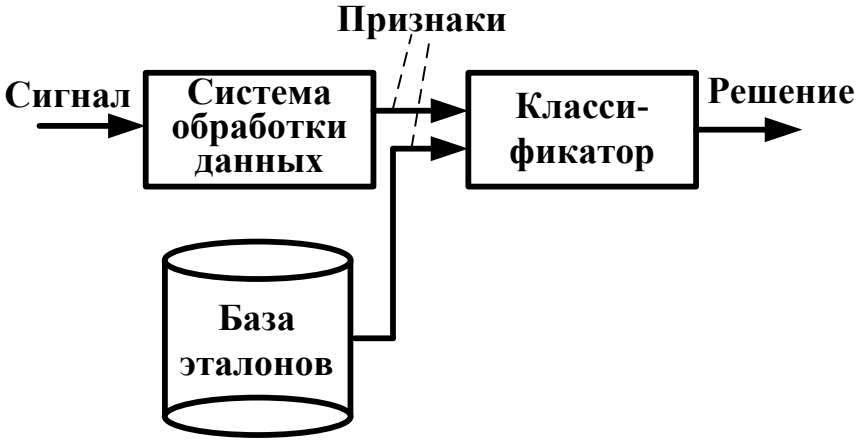

Рис. 3. Классификация принятого сигнала

3. Вычислительный эксперимент

3.1. Описание тестовых данных

В вычислительном эксперименте использовались звуковые записи следующих слов: Вперёд, Лево, Назад, Право, Стоп. Выполнено 50 записей для каждого слова. Вычисления и запись базы данных выполнялись на следующей конфигурации: процессор – Intel Core i5-2410M, объём оперативной памяти 8 Гб. Вычислительный эксперимент заключается в проверке точности работы описанного метода распознавания.

Таблица 1

Точность классификации в зависимости от числа дихотомий в процентах

	SVM. 1-1		SVM. 1-N	
	нормализованные	ненормализованные	нормализованные	ненормализованные
1	2	3	4	5
1 / 1	88	86	95	96
1 / 2	89	85	96	96

Окончание табл. 1

1	2	3	4	5
1 / 4	85	81	95	91
1 / 8	83	78	94	93
2 / 1	91	93	97	97
2 / 2	91	92	98	97
2 / 4	91	91	97	97
2 / 8	89	89	96	95
4 / 1	90	92	96	94
4 / 2	90	93	96	95
4 / 4	90	91	96	95
4 / 8	88	92	96	94

Выводы

Проведение предварительной нормализации значений повышает точность класификации в обоих подходах, но это увеличивает вычислительные затраты конечного алгритма. Подход «один против всех» позволяет получить более высокие показатели при разспознование. Как и подход «один против всех», так и «один против одного» дают достаточно высокие показатели, что позволяет выбирать при реализации наиболее подходящий по доступным вычислительным ресурсам.

Заключение

В работе рассматривается метод распознавания речевых команд в условиях априорной неопределенности в задачах управления робототехнической системой с позиции активного восприятия. Предлагается несколько вариантов классификаторов. Приводятся результаты вычислительного эксперимента.

Библиографический список

1. **Утробин, В. А.** Элементы теории активного восприятия изображений // Труды НГТУ им. Р.Е. Алексеева. 2010. Т. 81. № 2. С. 61–69.
2. Распознавание образов: состояние и перспективы / К. Верхаген [и др.]. – М.: Радио и связь, 1985. – 104 с.
3. **Загоруйко, Н. Г.** Методы распознавания и их применение / Н.Г. Загоруйко. – М.: Сов. радио, 1972. – 208 с.
4. **O'Shaughnessy, D.** Acoustic Analysis for Automatic Speech Recognition // Proceedings of the IEEE. – 2013. V. 101. N. 5. P. 1038–1053.
5. **Карасиков, М.Е.** Поиск эффективных методов снижения размерности при решении задач многоклассовой классификации путем её сведения к решению бинарных задач М.Е. Карасиков, Ю.В. Максимов // Машинное обучение и анализ данных. 2014. Т. 1. № 9. С. 1273–1290.
6. **Saon, G.** Large-Vocabulary Continuous Speech Recognition Systems: A Look at Some Recent Advances / G. Saon, J.-T. Chien // IEEE Signal Processing Magazine. 2012. V. 29. N. 6. P. 18–33.
7. **Котомин, А. В.** Распознавание речевых команд с использованием сверточных нейронных сетей // Наукоёмкие информационные технологии SIT-2012: труды молодежной конф. – Переславль-Залесский, 2012. С. 17–28.
8. **Котомин, А. В.** Предобработка звукового сигнала в системе распознавания речевых команд // Наукоёмкие информационные технологии SIT-2011: труды XV молодежной конф. – Переславль-Залесский, 2011. С. 25–38.

Дата поступления
в редакцию 02.07.2015

M. O. Derbasov[2], A. A. Laptev[3], A. A. Filyakov[1], V. E. Gai[1]

VOICE CONTROL OF ROBOTS FROM THE STANDPOINT OF THE THEORY OF ACTIVE PERCEPTION

Nizhny Novgorod state technical university n.a. R.E. Alexeev[1],
CJSC Intel[2],
Nizhny Novgorod radio engineering College[3]

This abstract related to a description of the method for recognition of voice commands in conditions of a priori uncertainty in control problems of robots. In contrast to a signal samples methods this method represents coarse-to-fine conception of signal analyze described in active perception theory. Provides results of computing experiment for confirming the efficiency of this method.

Proposed to implement two stages of recognition: preprocessing phase and phase calculation signs. At the stage of pre-treatment is performed the integration signal. At a stage of calculation of signs – the algebra of groups and operation of a dichotomy is used(the one-dimensional histogram of the closed groups). Dimension of system of signs for one sample are 4·840 elements. At the stage of classification used a support vector machine. Two approaches to creation of the multiclass qualifier are considered: «one-against-one» and «one - against – all».

Performed testing the proposed method based on cross-checking. The accuracy of the classification on a database of the 5 commands (50 realizations recorded each word) is 98%. The results can be used in the creation of methods of continuous speech recognition. The developed system of signs can also be used in other tasks classification signals.

*Key words***:** speech command recognition, theory of active perception.