

ОБ ИСПОЛЬЗОВАНИИ ПСИХОЛОГИЧЕСКИХ ДАННЫХ
ПРИ АВТОМАТИЧЕСКОМ РАСПОЗНАВАНИИ

В.И. Галунов

В литературе по автоматическому распознаванию часто говорится о выделении полезных признаков. Обычно под этими словами подразумевается получение некоторого — сокращенного по сравнению с входным — описания классифицируемых объектов, сохраняющего, однако, основные различительные свойства. Так, например, полезные признаки могут интерпретироваться как главные компоненты в фактор-анализе. Задача строгого построения сокращенного описания классифицируемых объектов в настоящее время достаточно трудно реализуема, так как требует очень больших затрат. В связи с этим всё чаще используются различные эвристические методы. В этом же плане целесообразно, на наш взгляд, использовать психологические данные. В настоящей статье критерий "полезности" вводится как понятие чисто психологическое. При этом возникает реальная возможность использования данных о человеческом восприятии при построении систем автоматического распознавания.

Рассмотрим следующую расширенную формулировку задачи распознавания образов. Имеется некоторое множество объектов $\{\omega_i\}$, на котором выделено подмножество $\{\omega_i\}_0 \in \{\omega_i\}$. Имеется некоторое устройство (интерпретатор), которое осуществляет отображение $\{\omega_i\}$ на некоторое другое множество (образов, интерпретаций), $\{Q_i\} = I\{\omega_i\}$. Помимо этого, имеется еще одно устройство (измеритель), которое измеряет на $\{\omega_i\}$ некоторый набор параметров $(x_1, x_2, \dots, x_n) = \bar{x}$ и осуществляет тем самым отображение $\{\omega_i\}$ на $\{\bar{x}_i\}$, $M\{\omega_i\} = \{\bar{x}_i\}$. Задачей опознавания является нахождение по конечному числу заданных

сопряженных точек $Q_i \in \{Q_i\}_0 \equiv I \{\omega_i\}_0$ и $\bar{x}_i \in \{\bar{x}_i\}_0 \equiv M \{\omega_i\}_0$ такого отображения $\{\bar{x}_i\}_0$ на $\{Q_i\}_0$, $T\{\bar{x}_i\}_0 = \{Q_i\}_0$, что равенство $TM\{\omega_i\}_0 = I\{\omega_i\}_0$ будет выполняться для всех $\omega_i \in \{\omega_i\}_0$. В обычной постановке задачи распознавания образов на такую формулировку наложены ограничения. Предполагается, что $\{Q_i\}_0$ конечно, номинативно и $\{Q_i\}_0 \equiv \{Q_i\}$. Определим задачу выделения полезных признаков как другой частный случай приведенной выше общей формулировки. А именно, предположим, что $\{Q_i\}$ — метрическое пространство.

Следует отметить, что имеется большой ряд задач, где основной целью является не распознавание образов, а именно выделение полезных признаков в смысле данного выше определения. Они характеризуются прежде всего тем, что множество $\{Q_i\}$ бесконечно, и, следовательно, на первый план выдвигается задача не абсолютной классификации, а относительного описания. В качестве примеров можно привести распознавание почерков или голосов, идентификацию спектрограмм и сейсмограмм и т.п. Наиболее же интересным примером является, вероятно, распознавание речи.

В самом общем виде задачу автоматического распознавания речи можно сформулировать следующим образом: необходимо построить устройство, понимающее речь аналогично тому, как это делает человек. Однако такая задача является слишком сложной и подразумевает умение понимать смысл машиной. Последнее попадает уже скорее в сферу исследований по автоматическому переводу. Можно сузить задачу и считать, что нашей целью является создание такого описания речевого сигнала, что оно сохраняет смысловое содержание, минимизируя поток информации. Следует сразу отметить, что так как мы не определили понятия "смысл", то единственным образцом здесь нам может служить человек. Обычно считается, что подобное описание можно осуществить по следующей схеме. Речевой сигнал разбивается на сегменты, относительно которых считается, что они могут распознаваться независимо друг от друга в предположении классической постановки задачи с номинативным характером $\{Q_i\}$. Элементами при этом является фонемная система данного языка.

Предположение о посегментном характере фонемного кода речи вытекает из неправильного истолкования психологических данных. Исходной точкой при этом служит то, что человек может даже бессмысленные звуко сочетания записать в фонетическом виде. Возможность того, что фонемная последовательность может выде-

литься не посегментно, а по большему отрезку речи, не рассматривается. Если же обратиться к психологическим данным, то можно видеть, что данные большого ряда работ по исследованию восприятия речи (см., например, [1,2]) свидетельствуют о том, что категориальных позвучных решений человек не производит. На уровне же сегмента он осуществляет только более или менее подробное описание его. В работе Р.В. Петерса [3] приводятся данные о поведении человека при восприятии отдельных звуков. Оказывается, что это поведение очень просто описывается при помощи метрической модели (евклидово пространство с очень малой размерностью: 2 — 3).

Таким образом, в свете психологических данных представляется целесообразным формулировать задачу посегментного распознавания речи (которая обычно и подразумевается, когда говорят об автоматическом распознавании речи) в терминах выделения полезных признаков. В данном случае человеческое восприятие явилось для нас источником правильной постановки задачи. Однако выбранное нами определение полезных признаков дает возможность использования психологических данных и в тех случаях, когда решается задача распознавания образов номинативного типа. Обычно при построении распознающих систем человек используется только для задания классификации. Однако часто он является еще и устройством, ее реально производящим при помощи выбранного нами описания $\bar{X}$. Можно предположить, что при этом поведение человека организовано достаточно целесообразно..

Довольно общепринятой гипотезой считается, что до принятия категориального решения о поступившем сигнале человек описывает его при помощи некоторого набора субъективно полезных непрерывных признаков. Для того, чтобы иметь возможность выделить эту систему признаков, необходимо поставить человека в такие условия, когда его поведение отличается от обычного категориального вида. С этой целью можно использовать сравнительно хорошо разработанную методику построения психологических шкал [4], где ставится задача, подобная сформулированной нами для выделения полезных признаков. Аналогичным образом там по сопряженным точкам ищется отображение физических характеристик стимула на психологическое пространство субъективных параметров, которое проявляется как некоторая структура внешнего поведения. Таким образом, мы получаем возможность искусственного сведения задачи распознавания к задаче выделения полезных признаков. Решение этой последней позволяет перейти от

практически случайного инструментального описания $\bar{X}$ к системе в определенном смысле более хороших признаков. Основным достоинством нового описания является значительное сужение потока информации.

Сведение задачи распознавания образов к выделению полезных признаков может дать определенные преимущества и с чисто математической точки зрения. Так как психологическая модель $\{Q_i\}$ в принципе строится в произвольной метрической системе, то можно единообразно описывать $\{\bar{X}_i\}$ и $\{Q_i\}$ и, следовательно, рассматривать отображение метрического пространства в себя.

В заключение отметим, что та частная задача распознавания, которая здесь названа выделением полезных признаков, возникает в большинстве поведенческих задач. Это связано с тем, что вне психологического эксперимента со специальной инструкцией поведение человека или животного практически непрерывно. При этом возникает проблема обучения, которая сводится к построению методом проб и ошибок по обучающей последовательности некоторой шкалы, связывающей рецепторное описание окружающей среды с поведенчески полезным эффекторным. Таким образом, в рамках решения задачи выделения полезных признаков возникает возможность взаимного полезного сотрудничества исследователей, занимающихся автоматическим распознаванием и психологией.

В заключение автор считает своим приятным долгом выразить искреннюю благодарность Л.А. Чистович за стимулирующие дискуссии.

Л и т е р а т у р а

1. Miller G.A. Decision units in the perception of speech. IRE Trans, 1962, IT-8, 81.
2. Чистович Л.А. и др. Речь, артикуляция и восприятие. Издательство "Наука", 1965 г.
3. Peters R.W. Dimensions of Perception for Consonants. JASA, 1963, 35, 1985-1989.
4. Handbook of Math. Psychol, 1963, N-Y.

Поступила в редакцию
28.VI.1965