ТЕОРИЯ СКРЫТЫХ МАРКОВСКИХ МОДЕЛЕЙ И ЕЕ ПРИМЕНЕНИЕ ДЛЯ ОПТИЧЕСКОГО РАСПОЗНАВАНИЯ ПЕЧАТНЫХ СИМВОЛОВ. ОБЗОР

Цопкало Н.Н.

При наблюдении за каким-либо процессом, протекающим в реальном мире, первоначально регистрируется его проявление, с помощью датчиков воспринимаются сигналы, порождаемые этим процессом. Проявления процесса обычно подвержены шумам и искажениям, не позволяющим получить информацию об источнике сигнала и описать процесс, породивший этот сигнал. Для решения этой проблемы создаются математические модели сигнала, в частности — стохастические модели. Здесь мы рассмотрим один из типов стохастических моделей сигнала — скрытые марковские модели. В основе построения такой модели лежит допущение о том, что сигнал может быть описан случайным процессом, параметры которого могут быть определены конкретным способом. Основы этой теории были опубликованы в статьях Баума и его коллег в конце 60-х – начале 70-х годов, а в 70-х годах ее результаты были применены для распознавания речи. До середины 80-х годов теория не была широко известна, поскольку была опубликована в математических журналах, которые не читались инженерами, работающими в области распознавания, а описания первых применений не содержали достаточно вводно-методического материала, что мешало понять результаты теории и правильно применить ее в задачах распознавания. Популярность теории значительно возросла лишь после выхода нескольких обзорных статей, содержащих детальную информацию, особенно статьи Рабинера [1], на которую ссылаются практически все статьи об использовании СММ. В 90-х годах СММ начинают применяться для распознавания печатного и рукописного текстов. Теория получила развитие для случая, когда необходимо описывать двумерный сигнал — изображение, такой вид моделей называют псевдо-двумерные скрытые марковские модели и применяют для инвариантного распознавания символов различных шрифтов, отличающихся нелинейными искажениями формы, а также для поиска лиц на фотоснимках.

Рассмотрим основные понятия теории и три фундаментальные задачи, которые она ставит. Дискретной цепью Маркова первого порядка называется система, находящаяся в каждый дискретный момент времени t в одном из N состояний $s_1, s_2, ..., s_N$, обозначаемом q_t, причем вероятности перехода в следующий момент времени $t+1$ в каждое из N состояний определяются только состоянием q_t и не зависят от $q_{t-1}, q_{t-2}, ..., q_1$. Формально такая система полностью описывается матрицей A с элементами a_{ij}:

$$a_{ij} = P[q_t = s_j | q_{t-1} = s_i], i, j = 1, ..., N. \quad (1)$$

Такой случайный процесс можно назвать наблюдаемой марковской моделью, поскольку состояние системы как раз и является ее выходом, т.е. соответствует регистрируемому физическому событию. Эта модель является слишком ограниченной и не подходит для решения многих практических задач,

поэтому понятие марковской модели было расширено. В новой модели наблюдаемые события являются некоторой вероятностной функцией текущего состояния, т.о. основной процесс оказывается невидимым, скрытым. Приведем пример вероятностного процесса извлечения шаров из урн. Пусть имеется N урн, содержащих большое количество шаров, каждый из которых окрашен в один из M цветов. По некоторому вероятностному закону случайно выбирается урна и из нее произвольно извлекается шар, цвет которого записывается как первое наблюдение. Затем выбирается следующая урна по вероятностному правилу, зависящему от номера первой урны, из нее также извлекается шар, цвет которого — второе наблюдение, и т.д. Такой процесс описывается СММ, состояния которой соответствуют выбираемым урнам, а для каждого состояния определены свои вероятности появления каждого из цветов. Множество наблюдаемых символов (в данном примере — множество цветов) обозначается как $V = \{v_1, v_2, ..., v_M\}$, а распределение вероятностей их появления в состоянии s_j как $B = \{b_j(k)\}$, где

$$b_j(k) = P[v_k|s_j], j = 1, ..., N; k = 1, ..., M. \quad (2)$$

Начальное распределение вероятностей состояний (закон, по которому выбирается первая урна) обозначается через $\Pi = \pi_i$:

$$\pi_i = P[q_1 = s_i], i = 1, ..., N \quad (3)$$

Наблюдение в момент времени t обозначается $o_t, o_t \in V$. Для практического применения теории должны быть решены три следующие проблемы:

1) Какова вероятность появления последовательности наблюдений $O = o_1, o_2, ..., o_T$ для модели $\lambda = (A, B, \Pi)$?

2) При заданных λ и O, как выбрать цепочку состояний $Q = q_1, q_2, ..., q_T$ которая некоторым наилучшим образом соответствует имеющейся последовательности наблюдений O?

3) Каким образом нужно подстроить параметры модели $\lambda = (A, B, \Pi)$, чтобы вероятность $P[O|\lambda]$ была максимальной?

Существуют классические способы решения этих задач, вполне приемлемые по вычислительной сложности. Первая задача решается посредством вычисления так называемой прямой переменной α (алгоритм прямого-обратного хода). Это позволяет определить, насколько хорошо данная модель соответствует наблюдениям O, а если имеется несколько моделей, позволяет выбрать из них ту, которая подходит наилучшим образом. Так, если построены модели речевых сигналов для слов, решение задачи 1 позволяет подобрать ту модель, которая соответствует неизвестному произнесенному слову. Для решения второй задачи обычно применяется алгоритм Витерби, использующий динамическое программирование и вычисляющий наилучшую цепочку состояний имеющую максимальную вероятность $P[Q|O, \lambda]$, однако существуют и другие критерии. В третьей задаче применяется процедура переоценки Баума-Уэлча. Цель состоит в том, чтобы так оптимизировать параметры модели λ, чтобы она наилучшим образом соответствовала O, называемой обучающей последовательностью, таким образом, создается модель, наиболее близко описывающая наблюдаемый процесс или явление.

В реальных задачах наблюдаемая величина обычно представляет собой вектор. Например, в распознавании речи это вектор спектральных характеристик речевого сигнала в данный момент времени. Для построения дискретной модели необходимо различать конечное число символов наблюдения, поэтому все возможные наблюдаемые векторы должны пройти процедуру кластеризации и быть сгруппированы в компактные области по некоторому критерию близости, а векторы, относящиеся к одной области будут обозначать один и тот же символ наблюдения.

Как мы видим, для построения СММ необходимо, чтобы наблюдаемая величина была функцией одной независимой переменной. При распознавании печатного текста имеется изображение целой страницы, и нет однозначного способа определения такой независимой переменной и вектора наблюдений. Обычно рассматривается сегментированная строка текста, в качестве независимой переменной выступает x-координата, а векторы наблюдений вычисляются из узкой вертикальной полосы пикселов с этой координатой. Рассмотрим особенности метода распознавания, предложенного в [2] для случая факсимильного текста. Основная идея состоит в том, что для символов алфавита строится по две модели, одна — на основе векторов наблюдений, вычисленных по столбцам пикселов изображения выделенного символа, а в другой векторы извлекаются из строк пикселов. Вычисленные степени соответствия неизвестного изображения символа этим двум моделям являются независимыми критериями, объединение которых повышает вероятность правильного распознавания. В интегрированную оценку включаются также степени соответствия ширины и высоты распознаваемого символа по отношению к эталонному. Вектор наблюдений представляет собой бинарный вектор, соответствующий ряду пикселов, центр масс которого приведен к начальному положению посредством сдвига. Такая нормализация уменьшает разнообразие наблюдений и упрощает процедуру кластеризации. Положение центра масс данного столбца пикселов, вычисленное относительно среднего положения трех предыдущих столбцов также выступает в качестве наблюдения. Создается лево-правая модель Бакиса, обычная для задач распознавания речи и текста. Ее особенность в том, что индекс состояния системы в следующий момент времени не убывает, но здесь наложено дополнительное ограничение: индекс не может увеличиваться более чем на 1. Матрица переходных вероятностей такой модели имеет ненулевые элементы только на главной и на находящейся над ней диагоналях. Причина использования такой модели в том, что в процессе обучения символ разбивается на зоны, в пределах которых векторы наблюдений имеют близкие значения, каждая из этих зон соответствует одному состоянию.

Рассмотрим столбцы пикселов, из которых состоит символ "m". Они образуют пять основных областей: три области с длинными участками черных пикселов, разделенные двумя областями с короткими участками. Для модели "m" с 5 состояниями, в пределах одной области длится соответствующее состояние, в котором высокие вероятности соответствуют символам наблюдений, отражающим характерный шаблон пикселов и низкие вероятности — остальным. Чтобы вычисленная степень соответствия неизвестной последо-

вательности была высокой, в ней должны присутствовать все эти области с характерными символами наблюдений, причем небольшие независимые изменения ширины областей не ухудшат эту оценку существенно.

И в задачах распознавания речи, и при распознавании текста существует сходная проблема сегментации. Как определить границы, разделяющие распознаваемые единицы: слова или фонемы в слитной речи и символы в тексте? Слипшиеся и распавшиеся символы на изображении слова представляют большую сложность для распознавания, а такая ситуация является типичной при факсимильной передаче. СММ, составленные по вектор-столбцам символов позволяют решить эту задачу с помощью так называемого алгоритма построения уровней. В результате на изображении слова выделяются отдельные символы, затем производится повторная оценка, используя модели вектор-строк.

Для повышения качества распознавания слов в работе [2] используются переходные вероятности букв английского языка и словарь. Эта информация использовалась двумя разными способами: в процессе работы алгоритма построения уровня и после сегментации и повторного оценивания символов. Второй вариант показал лучшие результаты, используя при коррекции алгоритм Витерби со словарем. Проведенные эксперименты показали, что комбинирование описанных методов позволяет снизить количество ошибочно распознанных слов более чем вдвое по сравнению с коммерческой программой OmniPage Pro, при работе с факсимильными сообщениями. При размере шрифта 14 пт. ошибочно распознанными оказались 14,8% слов.

Список литературы

[1]. *Рабинер Л.Р.* Скрытые марковские модели и их применение в избранных приложениях при распознавании речи: Обзор//ТИИЭР, т.77, N2, февраль 1989 — с. 86–120.

[2]. *Elms A.J., Procter S., Illingworth J.* The advantage of using an HMM-based approach for faxed word recognition// International Journal on Document Analysis and Recognition (1998) 1: 18–36