Статистическое представление изображений выборками случайных отсчетов фиксированного размера

В.Е. Анциперов[1]

[1]Институт радиотехники и электроники им. В.А. Котельникова РАН, Моховая 11/7, Москва, Россия, 125009

Аннотация

В работе обосновывается способ представления изображений выборками случайных фотоотсчетов и приводится статистическое описание подобных представлений. Предлагаемый способ опирается на вытекающую из фундаментальных фактов теории взаимодействия излучения с веществом концепцию идеального изображения. С целью сокращения размеров представления идеальное изображение редуцировано к выборкам отсчетов фиксированного (контролируемого) размера. Показано, что статистическое распределение последних факторизуется в произведение распределений отдельных отсчетов, аналогично модели наивного байесова подхода, это позволяет использовать для работы с предложенными представлениями богатый арсенал средств машинного (байесова) обучения.

Ключевые слова

вариационные автокодировщики, выборки случайных отсчетов, машинное обучение

1. Введение

Работа посвящена обоснованию и обсуждению одного способа представления данных для анализа широкого класса изображений. Предлагаемый способ ориентирован на обработку (распознавание, выделение характеристик, детектирование особенностей и т.д.) изображений с помощью вариационных автокодировщиков (VAE – variational autoencoders [1]).

Автокодировщики – это простые самообучаемые схемы, предназначенные для преобразования входных данных в подобные же им выходные с минимально возможным искажением, но с возможно большем сжатием промежуточного представления. Впервые автокодировщики появились в 1980-х годах, однако, относительно недавно вновь привлекли к себе внимание в связи с системами «глубокой архитектуры», где они самостоятельно (unsupervised) обучаются снизу вверх, после чего следует фаза контролируемого (supervised) обучения для верхнего уровня и точной настройки всей архитектуры.

Значительного улучшения характеристик автокодировщиков удалось достигнуть в последнее время в рамках модели вариационных автокодировщиков (VAE) [1]. Согласно байесовскому подходу, VAE относятся к классу порождающих (generative) моделей, которые рассматривают входные данные как случайные, порожденные некоторым вероятностным распределением. Как правило, порождающее распределение задается (с точностью до некоторых неизвестных параметров θ) посредством совместного распределения $p_\theta(X,Z)$, где X обозначает набор входных (наблюдаемых) данных, а Z – множество скрытых (ненаблюдаемых) переменных. Действительно, если для Z имеется некоторая оценка $\hat{Z}$, то новые данные $Y=X$, рассматриваемые как данные на выходе автокодировщика, порождаются (генерируются) с помощью распределения $p_\theta\big(Y|\hat{Z}\big)=p_\theta(Y,\hat{Z})/p_\theta(\hat{Z})$. Данная операция, называемая в контексте VAE декодированием, помимо оценок $\hat{Z}$ требует также задания значений параметров θ. Принципиально, эти величины могут быть найдены по входным данным X на предшествующем этапе кодирования по (апостериорным) распределениям $p_\theta(Z|X)=p_\theta(X,Z)/p_\theta(X)$, например, по схеме EM–алгоритма. Однако, ввиду того, что зависимость

$p_\theta(Z|X)$ от θ и Z часто оказывается очень сложной, для VAE кодирования обычно, по аналогии с вариационным байесовым подходом (variational Bayesian inference), используется приближенное апостериорное распределение $q_\varphi(Z|X)$ из более простого параметрического семейства (детали см. в [1]).

2. Представление изображений выборками случайных отсчетов

Как следует из вышеизложенного, задание модели кодирования (encoder model) $q_\varphi(Z|X)$ является для VAE принципиальным моментом, поскольку именно она в первую очередь определяет как характеристики автокодировщика, так и эффективность его реализации. При этом, очевидно, выбор апостериорного распределений $q_\varphi(Z|X)$ существенно зависит, помимо прочего, от вида представления X входных данных для соответствующей предметной области.

В течение последнего времени нами была проведена определенная работа по поиску подходящих представлений для изображений – вначале для слабоинтенсивных, а затем и для более широкого класса [2]. Мы исходили из концепции идеального изображения как совокупности фотоотсчетов $X = \{\vec{x}_i\}$ – физического результата регистрации потока квантов, падающих на чувствительную поверхность идеального устройства визуализации. Последнее представляет собой массив (матрицу) идеальных точечных детекторов, прототипами его являются сетчатка глаза животных и человека, фотопластинки с покрытием, содержащим светочувствительные зерна кристаллов галогенида серебра, светочувствительные КМОП-матрицы однофотонных фотодиодов.

Известно (см. обсуждение и дальнейшие ссылки в [2]), что представление изображений отсчётами статистически описывается как распределение вероятностей n точек $\{\vec{x}_i\}$ некоторого двумерного неоднородного точечного пуассоновского процесса с функцией интенсивности пропорциональной интенсивности падающего излучения. Однако, в отличие от случая слабых интенсивностей, в случае классических изображений $n \to \infty$ (10^8 и более), в результате чего работа непосредственно с представлением $X = \{\vec{x}_i\}$ представляется мало реалистичной.

Предложенное нами решение проблемы представления изображений аналогично упомянутому пуассоновскому, но изначально задается фиксированным размером $k \ll \bar{n}$. Именно, рассматривая представление идеального изображения $X = \{\vec{x}_i\}$ как генеральную совокупность случайных точек и осуществляя из нее случайную выборку в k отсчетов $X_k = \{\vec{x}_j\}$, мы объявляем ее представлением изображения. Нами показано, что в тех же предположениях, что использованы при выводе статистики пуассоновского представления, статистика выборок фиксированного (не случайного) размера k задается плотностями распределений вида:

$$\rho(\vec{x}_1, \dots, \vec{x}_k) = \prod_{j=1}^{k} \nu(\vec{x}_j), \qquad \nu(\vec{x}) = I(\vec{x}) \Big/ \int_\Omega I(\vec{x}) d\vec{x}\,,$$

где $I(\vec{x})$ – интенсивность излучения, падающего на чувствительную поверхность Ω устройства.

Отметим, что концепция идеального изображения понадобилась нам только для обоснования статистического описания $\rho(X_k)$ представления. Практическая же реализация представления $X_k = \{\vec{x}_j\}$, $\vec{x}_j \in \Omega$, состоит в формировании k независимых, одинаково распределенных (iid) случайных отсчетов $\vec{x}_j$, связанных с нормированной интенсивностью $\nu(\vec{x})$. В случае, когда выборка формируется в процессе регистрации излучения, отсчеты отбираются из потока фотоэлектронов по мере их эмиссии, пока их число не достигнет величины k. Если же изображение уже задано в виде набора пикселей, по ним можно восстановить (reconstruct) приближенные значения $I(\vec{x})$ в точках Ω. Нормируя это приближение, можно получить аппроксимацию плотности распределения $\nu(\vec{x})$, после чего остается произвести k раз процедуру семплирования. Отметим в этой связи, что в области машинного обучения имеется богатый арсенал методов сэмплирования, объединенный общим названием методов Монте-Карло.

3. Литература

[1] Kingma, D.P. An Introduction to Variational Autoencoders / D.P. Kingma, M. Welling // Foundations and Trends in Machine Learning – 2019. – Vol. 12(4). – P. 307-392. DOI: 10.1561/2200000056.

[2] Antsiperov, V. Maximum Similarity Method for Image Mining / V. Antsiperov // Pattern Recognition. ICPR International Workshops and Challenges. Lecture Notes in Computer Science. – 2021. – Vol. 12665. – P. 301-313.