

Генеративная модель автокодировщиков изображений на основе рецептивных полей

В.Е. Агциперов

Институт радиотехники и электроники им. В.А. Котельникова РАН

Москва, Россия

antciperov@cplire.ru

Аннотация—В работе обсуждается новая концепция автокодировщиков, ориентированных на кодирование цифровых изображений. Проблема синтеза подобных автокодировщиков формализуется в рамках подхода машинного обучения, понимаемого здесь как обучение (без учителя) по данным самого же изображения. Особенностью рассматриваемого подхода является специальное представление изображений с помощью выборок отсчетов контролируемого размера –выборочных представлений. Основываясь на специфике представления формулируется генеративная (порождающая) модель автокодировщиков, которая конкретизируется затем до вероятностной параметрической модели отсчетов в виде смеси компонент экспоненциального скошенного семейства. По аналогии с концепцией рецептивных полей компоненты смеси выбираются в виде финитных вероятностных распределений, что обеспечивает синтез быстрых рекуррентных процедур кодирования. Обсуждается связь синтезированных автокодировщиков со схемами мягкой и жесткой брегмановской кластеризации, классическим алгоритмом K–средних и алгоритмом LBG векторного квантования.

Ключевые слова— машинное обучение, генеративная модель, автокодировщики, метод максимального правдоподобия, смеси экспертов, рецептивные поля.

1. ВВЕДЕНИЕ

На сегодняшний день методы машинного обучения, ориентированные на нейро-сетевые структуры глубокого обучения, достигли впечатляющих результатов, их широкое распространение воспринимается как старт эпохи искусственного интеллекта (ИИ) [1]. Среди наиболее впечатляющих применений следует отметить распознавание слитной речи, перевод текста между языками, распознавание рукописей, распознавание лиц на изображениях и т.д. Отметим, что реализованный в этих приложениях уровень ИИ на сегодняшний день вполне сопоставим с интеллектом живого человека. Профессиональное ИТ–сообщество, разделяя всеобщее воодушевление успехами ИИ, вместе с тем все чаще выражают некоторую озабоченность сложившимся положением дел. Это связано с тем, что до сих пор отсутствует ясное понимание того, почему ряд подходов, зачастую эвристических, оказался на практике существенно более эффективным, чем методы машинного обучения, разработанные до начала столетия.

Частичный ответ на этот вопрос может быть связан с главной особенностью методов машинного обучения – их ориентированностью на данные. Отсюда основные различия в характеристиках методов обусловлены, с одной стороны, спецификацией используемых обучающих данных, а, с другой – с выбором модели зависимости принимаемых по данным решений. Под спецификацией подразумевается наличие или отсутствие

в обучающих данных явной разметки (меток принимаемых решений). Это подразделяет методы машинного обучения на обучение контролируемое (с учителем) и неконтролируемое (без учителя). С точки же зрения используемой (статистической) модели, связывающей данные с разметкой, методы машинного обучения подразделяются на дискриминантное и генеративное, в зависимости от того, основана ли модель разметки на условном по отношению к данным вероятностном распределении меток или на совместном их распределении [2]. Машинное обучение знает примеры различных сочетаний обеих дихотомий – обучения с учителем / без учителя и дискриминантного / генеративного моделирования, что позволяет выделить в них, соответственно, четыре класса. Сравнений методов по отношению к приведенной классификации привело к тому, что суммарно, по совокупности характеристик наиболее успешными оказались подходы основанные на генеративных моделях обучения без учителя. К этим подходам, в частности, относятся генеративно-состязательные сети (GAN) [3], вариационные автокодировщики (VAE) [4], глубокие сети доверия (DBN) [5] и т.д.

2. МОДЕЛИРОВАНИЕ НА ОСНОВЕ ПРЕДСТАВЛЕНИЙ

Причина успешности генеративных моделей обучения без учителя не вполне ясна, однако, часто в качестве аргумента приводится тот факт, что они более последовательно моделируют механизмы естественного (человеческого / высших млекопитающих) интеллекта. Действительно, по мере развития перечисленных выше моделей [3-5], каждый новый этап разработки добавлял им новые функции / элементы, имитирующие, например, иерархическую архитектуру областей коры головного мозга, глубокое обучение с подкреплением, кратковременную память в корковых сетях, долговременную память о событиях (более подробно см. [6]). При этом, поскольку детали реализации информационных процессов в коре головного мозга ясны не до конца, соответствующие функции / элементы приходится моделировать на базе компьютерных структур *представления данных*. Эти представления необходимо должны быть достаточно гибкими и масштабируемыми, чтобы допускать многоуровневость, инвариантность, иерархические зависимости и пр., позволяя представлениям на более высоком уровне быть более абстрактными и инвариантными к вариациям, которые обычно присутствуют в обучающем (низкоуровневом) представлении, избегая при этом по возможности потери существенной информации [7]. В этой связи возникает интересный обратный вопрос – насколько выбор тех или иных представлений входных – промежуточных – выходных данных предопределяет синтез ожидаемых функций модели?

В данной работе, ориентируясь на генеративную модель обучения без учителя [2], применительно к задачам (авто)кодирования изображений, изложены результаты синтеза нового рекуррентного метода машинного обучения на основе представлений данных, сформированных по самим же изображениям.

3. АВТОКОДИРОВЩИК НА ОСНОВЕ ПРЕДСТАВЛЕНИЙ

В качестве представления входных данных использованы выборки контролируемого размера в k отсчетов, сэмпированных из кодируемых изображений (выборочные представления), детальное обсуждение которых можно найти в [8,9]. Коротко, выборочные представления состоят из k -наборов координат случайных отсчетов $X_k = (x_1, \dots, x_k)$, распределение которых по области изображения Ω задается плотностью распределения:

$$\rho(X_k|I(\vec{x})) = \prod_{j=1}^k \rho(\vec{x}_j|I(\vec{x})), \rho(\vec{x}_j|I(\vec{x})) = \frac{I(\vec{x})}{\iint_{\Omega} I(\vec{x}) ds}. \quad (1)$$

где $I(\mathbf{x})$ – интенсивность формирующего изображение излучения, кратная, например, его битовой карте.

В рамках генеративной модели предлагаемого метода распределение вероятностей отдельного отсчета $\rho(x_j|I(\mathbf{x}))$ задается с помощью плотности распределения вероятностей из некоторого параметрического семейства смесей экспертов (смеси компонент) [10]:

$$\rho(\vec{x}|I(\vec{x})) = \sum_{z=1}^K w_z \eta(\vec{x} - \vec{\mu}_z; \vec{v}_z), \eta(\vec{x}; \vec{v}) = \exp\{\vec{v}\vec{x}^T - A(\vec{v})\} \rho_0(\vec{x}). \quad (2)$$

где K – число экспертов, $\{w_z\}$ – рейтинги экспертов, $\{v_z\}$ – экспертные оценки, задающие вид $\eta(\mathbf{x} - \mu_z; v_z)$ – локальных экспертных плотностей распределений \mathbf{x} . Каждое экспертное распределение задается как сдвинутый в точку $\mu_z \in \Omega$ элемент параметрического экспоненциально-скошенного семейства плотностей $\{\eta(\mathbf{x} - \mu_z; v)\}$. Выбранный вид (2) генеративной модели мотивирован концепцией рецептивных полей зрительной системы [11]. Именно, если под экспертами понимать ганглиозные клетки, а под μ_z расположения контролируемых ими областей рецепторов сетчатки Ω , то $\rho(x_j|I(\mathbf{x}))$ задает модель зрительной нейронной активности. Данная ассоциация подсказывает вид $\rho_0(\vec{x})$ как симметричной плотности с компактным носителем Δ . При этом носителями экспертных плотностей будут Δ_z – смещенные в точки μ_z компакты Δ , задающие систему (перекрывающихся) рецептивных полей.

Рассматривая параметры экспертов $\Theta = \{w_z, v_z\}$ в (2) как кодировку изображения (интенсивности $I(\mathbf{x})$), можно найти их оптимальные (наиболее вероятные) значения для представления X_k . Используя, например, метод максимального правдоподобия в отношении распределения отсчетов (1), можно получить, аналогично тому, как это делается при выводе рекуррентного EM-алгоритма [12], следующий метод оптимального кодирования:

Шаг I:

$$\rho_j^{(n+1)}(z) = \frac{w_z^{(n)} \exp\{-B_{A^*}(\vec{x}_j - \vec{\mu}_z, \vec{\zeta}_z^{(n)})\}}{\sum_{i \in \delta_{\vec{x}_j}} w_i^{(n)} \exp\{-B_{A^*}(\vec{x}_j - \vec{\mu}_i, \vec{\zeta}_i^{(n)})\}} \quad (3)$$

Шаг II:

$$w_z^{(n+1)} = \frac{1}{k} \sum_{\vec{x}_j \in \Delta_z} \rho_j^{(n+1)}(z), \quad (4)$$

$$\vec{\zeta}_z^{(n+1)} = \frac{\sum_{\vec{x}_j \in \Delta_z} \rho_j^{(n+1)}(z) (\vec{x}_j - \vec{\mu}_z)}{\sum_{\vec{x}_j \in \Delta_z} \rho_j^{(n+1)}(z)}.$$

где $n = 1, 2, \dots$ – счетчик итераций, индексы $i \in \delta_{\vec{x}_j}$ экспертов задаются рецептивным окружением отсчета x_j , ζ_z – связанные преобразованием Лежандра с оценками v_z параметры средних, $B_{A^*}(\mathbf{x} - \mu_z; \zeta_z)$ – дивергенция Брегмана [13] для сопряженной $A(v)$ (2) по Лежандру функции $A^*(\zeta)$.

Следует отметить, что метод кодирования (3,4) с точностью до специфики, связанной с рецептивными полями, ограничивающей в (3) суммирование до δ_{x_j} , а в (4) до Δ_z , совпадает с алгоритмом мягкой брегмановской кластеризации, описанной в [13]. Более того, как показано там же, при переходе к жесткой брегмановской кластеризации и при соответствующем выборе функции $A^*(\zeta)$ получается целый спектр известных в области машинного обучения алгоритмов: классический алгоритм K -средних (K -means) [14], алгоритм LBG векторного квантования [15] и алгоритм теоретико-информационной кластеризации [13].

ЛИТЕРАТУРА

- [1] Sejnowski, T.J. The unreasonable effectiveness of deep learning in artificial intelligence. // Proceedings of the National Academy of Sciences. – 2020. – Vol. 117(48). – P. 30033-30038. DOI: 10.1073/pnas.1907373117.
- [2] Bishop, C.M. Generative or Discriminative? Getting the Best of Both Worlds / C.M. Bishop, J. Lasserre. // Bayesian Statistics 8, International Society for Bayesian Analysis. – Oxford University Press. – 2007. – P. 3-24.
- [3] Goodfellow, I. Generative Adversarial Networks. / I. Goodfellow // Communications of the ACM. – 2020. – Vol. 63(11). – P. 139-144. DOI: 10.1145/3422622.
- [4] Kingma, D.P. Auto-Encoding Variational Bayes. / D.P. Kingma, M. Welling // CoRR abs/1312.6114, 2014.
- [5] Hinton, G.E. A Fast-Learning Algorithm for Deep Belief Nets. / G.E. Hinton, S. Osindero, Y.-W. Teh // Neural computation. – 2006. – Vol. 18(7). – P. 1527-1554. DOI: 10.1162/neco.2006.18.7.1527.
- [6] Hassabis, D. Neuroscience-Inspired Artificial Intelligence / D. Hassabis, D. Kumaran, C. Summerfield, M. Botvinick // Neuron. – 2017. – Vol. 95(2). – P. 245-258. DOI:10.1016/j.neuron.2017.06.011.
- [7] Bengio, Y. Deep learning of representations for unsupervised and transfer learning // Proceedings of the 2011 International Conference on Unsupervised and Transfer Learning workshop. – 2011. – Vol. 27. – P. 17-37.
- [8] Анциперов, В.Е. Статистическое представление изображений выборками случайных отсчетов фиксированного размера // Сб. трудов по материалам VII Межд. конф. и молодежной школы ИТНТ. – Самара: Самарский. нац. иссл. университет им. академика С.П. Королева. – 2021. – С. 24173.
- [9] Antsiperov, V.E. Image Coding by Count Sample, Motivated by the Mechanisms of Light Perception in the Visual System. / V.E. Antsiperov, V.A. Kershner. // Communications in Computer and Information Science. – 2022. – Vol 1534. DOI: 10.1007/978-3-030-96040-7_54.
- [10] Jacobs, R.A. Adaptive Mixtures of Local Experts / R.A. Jacobs, M.I. Jordan, S.J. Nowlan, G.E. Hinton // Neural Computation. – 1991. – Vol. 3(1). – P. 79-87. DOI: 10.1162/neco.1991.3.1.79.
- [11] Hyvärinen, A. Natural Image Statistics: Computational Imaging and Vision / A. Hyvärinen, J. Hurri, P.O. Hoyer. – London: Springer-Verlag, 2009. DOI: 10.1007/978-1-84882-491-1.
- [12] McLachlan, G.J. The EM Algorithm and Extensions / G.J. McLachlan, T. Krishnan. – New York: Wiley, 1997.
- [13] Banerjee, A. Clustering with Bregman Divergences / A. Banerjee, S. Merugu, I.S. Dhillon, J. Ghosh // Machine Learning Research. – 2005. – Vol. 6. – P. 1705-1749.
- [14] Lloyd, S. Least squares quantization in PCM // IEEE Transactions on Information Theory. – 1982. – Vol. 28(2). – P. 129-137. DOI: 10.1109/TIT.1982.1056489.
- [15] Linde, Y. An algorithm for vector quantizer design / Y. Linde, A. Buzo, R.M. Gray // IEEE Transactions on Communications. – 1980. – Vol. 28(1). – P. 84-95. DOI: 10.1109/TCOM.1980.1094577.