

Термодинамический формализм для исследования и конструирования алгоритмов и нейронных сетей

И.Ф. Ясинский, Ф.Н. Ясинский
ФГБОУВПО «Ивановский государственный энергетический университет имени В.И. Ленина»,
Иваново, Российская Федерация
E-mail: igor2266@yandex.ru

Авторское резюме

Состояние вопроса: А.Н. Колмогоров ввел понятие о вычислительной энтропии. После этого возник целый ряд работ (Тихомиров В.М., Витушкин А.Г., Петри Н.В.), в которых понятие энтропии было эффективно использовано для оценки сложности алгоритмов и программ.

Материалы и методы: Используются идеи и методы феноменологической и статистической термодинамики для оценки объема вычислений или размера нейронной сети.

Результаты: Выполнены вычисления с использованием, кроме энтропии, других термодинамических функций, а также определены три термодинамические начала, позволяющие исследовать устойчивость, упорядочивать параметры по их информационному весу, выполнять декомпозицию сложной системы, строить быстрые алгоритмы.

Выводы: Термодинамический формализм открывает новые возможности в создании оптимальных алгоритмов для моделирования сложных динамических систем и конструирования эффективных нейронных сетей, что может иметь очевидное приложение в энергетике.

Ключевые слова: вычислительная энтропия, быстрые алгоритмы, термодинамика вычислений.

Thermodynamic Formalism for Research and Construction of Algorithms and Neural Networks

I.F. Yasinskiy, F.N. Yasinskiy
Ivanovo State Power Engineering University, Ivanovo, Russian Federation
E-mail: igor2266@yandex.ru

Abstract

Background: A.Kolmogorov introduced the concept of computational entropy. After that a number of researches appeared (Tihomirov V.M, Vitushkin A.G., Petri N.V.), where the entropy was effectively used for the complexity evaluation of algorithms and programs.

Materials and methods: Ideas and methods of phenomenological and statistic thermodynamics are used to estimate the amount of calculations or volume of the neural network.

Results: Introduction of the other thermodynamic functions, besides entropy, and the definition of the three thermodynamic origins in the context of calculations allows to study stability, organize the parameters according their information weights, carry out the decomposition of the complex systems, construct the rapid algorithms.

Conclusion: Thermodynamic formalism opens up new possibilities in building of the optimal algorithms for modeling of the complex dynamic systems and construction of the effective neural networks that will find certain application in power engineering.

Key words: computational entropy, rapid algorithms, thermodynamics of calculations.

Вычислительная практика показывает, что вычислительный процесс и структуру нейронных сетей удобно исследовать, используя следующий термодинамический формализм.

Пусть имеет место итерационный вычислительный процесс:

$$x^{k+1} = F(x^k, \xi^k), \quad (1)$$

где $x^k \equiv (x_1^k, x_2^k, \dots, x_n^k)$ – точка в n -мерном фазовом пространстве X ; F – некий алгоритм; k – номер итерации; ξ^k – параметр учета шума.

Если процесс сходится, то

$$x_i^k \rightarrow x_i^*, k \rightarrow \infty, i = 1, 2, \dots, n, \quad (2)$$

где x^* – финальная точка, $x^* \in X$.

Возьмем в X множество фазовых точек с плотностью $\rho(x_1, \dots, x_n)$. Это множество можно рассматривать как некий «фазовый газ», который в процессе вычислений сжимается к точке x^* . Рассмотрим этот процесс с точки зрения термодинамики. Первые результаты в этом направлении принадлежат А.Н. Колмогорову и его группе. Они ввели понятие о вычислительной ε -энтропии и исследовали ее свойства [1–5]. Ниже вводим остальные термодинамические функции и формулируем термодинамические постулаты для вычислительного процесса.

Первое начало. Первое начало имеет вид

$$dN = T \cdot dl + p \cdot dn + \mu \cdot d\pi, \quad (3)$$

где N – число операций, затраченное на сжатие «фазового газа»; N – аналог внутренней энергии (обладает аддитивностью); I – информация, полученная в процессе вычислений.

Алгоритмическая температура

$$T = \frac{\partial N}{\partial I} \Big|_{n,\pi} \quad (4)$$

представляет число операций, затраченных на получение единицы информации.

Давление

$$p = \frac{\partial N}{\partial n} \Big|_{I,\pi} \quad (5)$$

является числом операций, потраченных на изменение размерности, если в процессе вычислений это имело место.

Слагаемое $\mu \cdot d\pi$ в (3) учитывает вероятностный характер вычислительного процесса (погрешности, использование случайных чисел), где

$$\mu = \frac{\partial N}{\partial \pi} \Big|_{I,n}, \quad (6)$$

π – вероятность правильного ответа, $0 < \pi < 1$. В ряде случаев определенное снижение π ведет к существенному сокращению числа операций N (методы Монте-Карло, случайный поиск).

Второе начало. Термодинамическая необратимость. Вычислительный процесс устойчив, если

$$dI > 0. \quad (7)$$

Опираясь на это неравенство, можно исследовать алгоритмы на устойчивость. Это значительно более эффективный нетрадиционный способ, он не требует линеаризации и применим в сложных нелинейных случаях. Убедительный пример приведен Ю.Н. Ватолиным [6]. В результате вычислений была получена новая информация:

$$\Delta I^{k,k+m} = H^k - H^{k+m},$$

где H^k, H^{k+m} – вычислительная энтропия на шагах k и $k+m$.

Третье начало. Недостижимость температурного нуля:

$$T = \frac{\partial N}{\partial I} \Big|_{n,\pi} > 0. \quad (9)$$

Весьма интересна величина, обратная алгоритмической температуре. Мы называем ее пропускной способностью алгоритма. Это извлеченная информация, отнесенная к числу затраченных операций:

$$C^{K_1, K_2} = \frac{\Delta I^{K_1, K_2}}{\Delta N^{K_1, K_2}}. \quad (10)$$

Вариационная задача о поиске оптимального распределения начальной плотности «фазового газа», минимизирующего число операций ΔN при заданной информации ΔI , приводит к распределению Л.Больцмана:

$$\rho^0(x_1, \dots, x_N) = \exp\left(-\frac{\Delta N}{T}\right) \cdot \left(\sum_{i=1}^n \exp\left(\frac{\Delta N}{T}\right)\right)^{-1}. \quad (11)$$

Из (11) элементарно получаются все известные соотношения статистической термодинамики, согласующиеся с изложенным феноменологическим подходом. Если согласно методу И. Ньютона линеаризовать локально отображение (1):

$$y_i^{k+1} = \sum_{j=1}^n A_{ij} y_j^k + \zeta_i^k, \quad (12)$$

$$y_i = x_i - x_i^*, \quad A_{ij} = \frac{\partial F_i}{\partial x_j},$$

и ввести информацию Хартли [1]:

$$\Delta I^{0,k} = -\ln \frac{V^k}{V^0}, \quad (13)$$

где V^0, V^k – начальный и конечный объемы в фазовом пространстве X , то можно показать, что

$$\Delta I^{0,k} = -\sum_{i=1}^n \ln \left(\lambda_i^k + \chi_i \frac{1 - \lambda_i^{k-1}}{1 - \lambda_i} \right), \quad (14)$$

где λ_i – собственные числа матрицы A ; χ_i – константы, пропорциональные уровню шума для частот λ_i .

В хорошо обусловленном случае, если принять

$$\lambda_s = \lambda; \quad \chi_s = \chi; \quad s = 1, 2, \dots, n,$$

а затраты на вычисления считать пропорциональными $k \cdot n(v + n)$, получим выражение

$$p(v + n) = IT, \quad (15)$$

которое можно назвать уравнением состояния алгоритма. В более сложных случаях уравнения состояния не будут такими простыми.

К итерационным вычислительным процессам применимо представление о термодинамических циклах. В качестве примера (см. таблицу) рассмотрим поиск минимума с помощью метода градиента:

$$x_i^{k+1} = x_i^k - \tau \frac{\partial Q}{\partial x_i} \Big|_{x^k}, \quad (16)$$

сопоставив ему цикл Карно в координатах (I, T) .

Пусть $0 < I_1 < I_2, 0 < T_1 < T_2$.

Участок	Температура начальная	Температура конечная	Информация начальная	Информация конечная
1	T_2	T_1	I_1	I_1
2	T_1	T_1	I_1	I_2
3	T_1	T_2	I_2	I_2
4	T_2	T_2	I_2	I_1

Участок 1. Вычисляются составляющие градиента, дающие направление «на цель». Растет пропускная способность, падает температура.

Участок 2. Движение к цели по полученному направлению. Падают энтропия, растет информация.

Участок 3. При новом положении поисковой точки вычисленные составляющие градиента уже не представляют ценности. Они отбрасываются. Падает пропускная способность и растет температура.

Участок 4. В новой точке принимается решение использовать тот же метод. Происходит возврат в исходную точку. На диаграмме (I, T) цикл замкнулся.

Описанный формализм позволяет, кроме исследования на устойчивость, находить оптимальные пути декомпозиции, выделять главные переменные, понижать размерность задачи, результативно рандомизировать вычисления.

Обратимся теперь к нейронным сетям. В этом случае вместо числа операций N вводится величина ε , пропорциональная рассеянию энергии нейронной сетью:

$$\varepsilon = b_n \cdot n + b_m \cdot m, \quad (17)$$

где n – число нейронов; m – число связей между ними; b_n, b_m – энергии, рассеиваемые нейронами и связями соответственно.

Выражения (3), (4), (5), (6) перепишутся в следующем виде:

$$d\varepsilon = T \cdot dl + p \cdot dn + \mu \cdot d\pi, \quad (18)$$

$$T = \left. \frac{\partial \varepsilon}{\partial l} \right|_{n, \pi}, \quad p = \left. \frac{\partial \varepsilon}{\partial n} \right|_{l, \pi}, \quad \mu = \left. \frac{\partial \varepsilon}{\partial \pi} \right|_{l, n, m}. \quad (19)$$

Если предположить, что все нейроны работают в областях быстрееших переходов, где локально допустима линеаризация, то можно воспользоваться выражением типа (14):

$$\Delta I^{o,k} = - \sum_{i=1}^n \ln \left(\lambda_i^k + \chi_i \frac{1 - \lambda_i^{k-1}}{1 - \lambda_i} \right). \quad (20)$$

Здесь k – порядковый номер нейронного слоя при прямом проходе (от входа к выходу). Опираясь на это, можно ставить задачи на выделение главных компонент, на упрощение, декомпозицию сети по наиболее слабым информационным связям, переход к новым переменным. Наличие слагаемого $\mu \cdot d\pi$ позволяет предположить, что рандомизация процессов в нейронной сети с использованием марковских алгоритмов может быть эффективной.

Компьютерное моделирование и построение быстрых алгоритмов актуальны для множества областей энергетики. Они используются при управлении ядерными реакторами,

прогнозировании процессов и управлении энергоресурсами.

Список литературы

1. Хинчин А.Я. Понятие энтропии в теории вероятностей // Успехи математических наук. – 1953. – Т. XI. – Вып. 3(55).
2. Колмогоров А.Н., Тихомиров В.М. ε -энтропия и ε -емкость множеств в функциональных пространствах // Успехи математических наук. – 1959. – 14. – Вып. 2(86). – С. 3–80.
3. Витушкин А.Г. Оценка сложности задачи табулирования. – М.: Физматиз, 1959.
4. Колмогоров А.Н. Различные подходы к оценке трудности приближенного задания и вычисления функций // Труды международного конгресса математиков. – Стокгольм, 1964.
5. Петри Н.В. Сложность алгоритмов и время их работы // ДАН СССР. – 1969. – 186. – С. 30–31.
6. Ватолин Ю.Н. О применении энтропийных оценок устойчивости // Численные методы механики сплошной среды. – 1974. – Т. 5. – № 2.
7. Бабенко К.И. Теоретические основы и конструирование численных алгоритмов задач математической физики. – М.: Наука, 1979. – 295 с.
8. Ахо А., Хопкрофт Дж., Ульман Дж. Построение и анализ вычислительных алгоритмов. – М.: Мир, 1979. – 535 с.
9. Трауб Дж., Вожняковский Х. Общая теория оптимальных алгоритмов. – М.: Мир, 1983. – 381 с.

References

1. Khinchin, A.Ya. Ponyatie entropii v teorii veroyatnostey [The concept of entropy in probability theory]. *Uspekhi matematicheskikh nauk*, 1953, vol. XI, issue 3(55).
2. Kolmogorov, A.N., Tikhomirov, V.M. *Uspekhi matematicheskikh nauk*, 1959, vol. 14, issue 2(86), pp. 3–80.
3. Vitushkin, A.G. *Otsenka slozhnosti zadachi tabulirovaniya* [Assessment of the tabulation problem]. Moscow, Fizmatiz, 1959.
4. Kolmogorov, A.N. Razlichnye podkhody k otsenke trudnosti priblizhennogo zadaniya i vychisleniya funktsiy [The different approaches to the difficulty estimation of the approximate setting and computing of functions], in *Trudy mezhdunarodnogo kongressa matematikov* [Proceedings of the International Congress of Mathematicians]. Stokgol'm, 1964.
5. Petri, N.V. *DAN USSR*, 186 (1969), pp. 30–31.
6. Vatolin, Yu.N. O primeneni entropiynykh otsenok ustoychivosti [On the application of the entropy bounds of stability], in *Chislennye metody mekhaniki sploshnoy sredy* [Numerical methods of continuum mechanics], 1974, vol. 5, no. 2.
7. Babenko, K.I. *Teoreticheskie osnovy i konstruirovaniye chislennykh algoritmov zadach matematicheskoy fiziki* [Theoretical foundations and construction of numerical algorithms for problems of mathematical physics]. Moscow, Nauka, 1979. 295 p.
8. Akho, A., Khopkroft, Dzh., Ul'man, Dzh. *Postroenie i analiz vychislitel'nykh algoritmov* [Design and Analysis of Computational Algorithms]. Moscow, Mir, 1979. 535 p.
9. Traub, Dzh., Vozh'nyakovskiy, Kh. *Obshchaya teoriya optimal'nykh algoritmov* [General Theory of Optimal Algorithms], Moscow, Mir, 1983. 381 p.

Ясинский Игорь Федорович,

ФГБОУВПО «Ивановский государственный энергетический университет имени В.И. Ленина»,
кандидат технических наук, доцент кафедры высокопроизводительных вычислительных систем,
e-mail: igor2266@yandex.ru

Ясинский Федор Николаевич,

ФГБОУВПО «Ивановский государственный энергетический университет имени В.И. Ленина»,
доктор физико-математических наук, профессор кафедры высокопроизводительных вычислительных систем,
телефон (4932) 26-98-29.