

МОТИВАЦИОННОЕ ПИСЬМО

Чернышев Алексей, 2013

Содержание

1	Вводные слова	3
2	Машинное обучение, основные проблемы	3
3	Reservoir computing	3
4	Нейронные сети, тренд развития	4
5	Новости с фронтов Neuroscience	5
6	Основная гипотеза	6
7	Задача и модель для проверки гипотезы	6
8	Заключение	7
	Список литературы	8

1 Вводные слова

Основная задача будущего исследования — создать унифицированный подход в использовании возможностей сложных динамических нейронных систем в проблемах обработки больших объемов данных и осуществления при помощи них классических задач машинного обучения таких как кластеризация, регрессия, классификация.

Динамические системы вдохновленные структурой коры головного мозга давно являются, как минимум, любопытным феноменом для учёных. Мозг является сложным многосистемным вычислительным устройством, который поражает своими вычислительными способностями и подчеркнуть хоть немного из его строения, научиться это использовать, является большим вызовом.

2 Машинное обучение, основные проблемы

Одной из классических задач машинного обучения является найти зависимость между скрытыми переменными (причиной) и следствием (данными). С такой задачей машинное обучение уже давно с успехом справляется. Можно взять в качестве примера наивный байесовский классификатор, в котором делается предположение о независимости признаков друг от друга (то есть скрытые переменные не взаимосвязаны причинно-следственной связью) и в итоге получить классификатор, который будет классифицировать объекты с той или иной погрешностью.

Основная проблема подобного подхода в том, что в реальных сложных задачах мы почти никогда не встретим случая, когда данные порождены абсолютно независимыми причинами.

Существуют графические вероятностные модели [1] в которых подобные зависимости вводятся в виде графа факторов. Также существует много способов обучать такие модели, но так или иначе, все они сталкиваются с невозможностью описать зависимость всех факторов друг от друга. Если, например, рассмотреть в качестве данных картинку 32×32 , в которой каждый пиксель принимает значение $\{0, 1\}$, то, чтобы описать зависимость всех факторов друг от друга, необходимо ввести 2^{1024} дополнительных факторов, не говоря о том, как сложно их обучить.

В 2003 году, В. Taskar предложил графическую модель которая может учитывать попарно связанные факторы [2], но тем не менее задача как рассматривать более высокоранговые связи внутри данных всё также остра.

3 Reservoir computing

Почти одновременно и с большой помпой вышли две работы Н. Jaeger [3] и W. Maass [4] в 2004 г., описывающие модели *Echo State Networks* и *Liquid State Machine* соотв., которые положили начало новому взгляду на использования рекуррентных сетей. Рекуррентные сети не без успеха использовались и до выхода работ, но рекуррентность в сетях порождает труднопреодолимые проблемы, как нестабильность и проблема обучения таких сетей [5].

Успех вышеуказанных работ заключается в уникальном по простоте подходе к динамическим системам (см. рис 1).

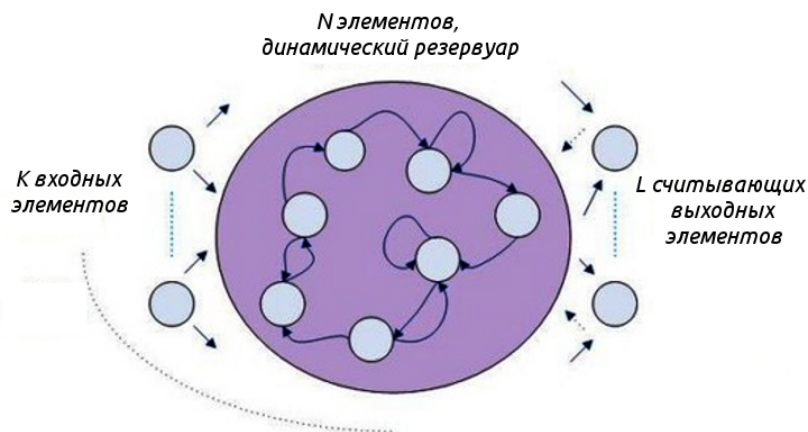


Рис. 1: схема Reservoir Computing

Основной принцип работы модели в том, что вычислительный резервуар, который представляет собой динамическую рекуррентную нейронную сеть, принимает уникальное состояние для каждого элемента данных, выходные же элементы по средством настраиваемых весов, обучаются на состояниях резервуара.

Методы обучения считывающих (*readout*) элементов на состояниях резервуара могут быть самыми примитивными – используя только решение системы нормальных уравнений для регрессии методом наименьших квадратов, уже можно добиться хороших предсказательных способностей [6].

Примечателен тот факт, что веса связей между элементами резервуара во время обучения остаются постоянными.

4 Нейронные сети, тренд развития

Первое поколение Искусственных Нейронных Сетей (ИНС) взяло только основные принципы работы биологических нейросетей. В них нейрон формализован как сумматор принимающий на вход сигналы от других нейронов и на основе искусственно введенной нелинейности решает передавать сигнал дальше или нет.

Вторым поколением ИНС можно назвать модели школы *Deep Learning*, которую основал J. Hinton с выходом работы [7]. Эти модели имеют стохастическую природу - нейрон передает или не передает сигнал в зависимости от вероятности выражаемой через коэффициенты модели, которые обучаются на данных без учителя.

Обучение без учителя является ключевым показателем интеллектуальности алгоритма. Интеллект сам по себе, как свойство живых организмов, представляет собой следствие самоорганизации природы в определенных условиях [8]. Руководствуясь подобными предпосылками, на базе школы *Deep Learning* выросло направление в *Machine Learning - Representation*

Learning [9].

Representation Learning собрал для себя следующие принципы, которые, по большей части, взяты именно из особенностей работы мозга:

- распределенность хранения (distributed representation);
- разреженность представления (sparse code);
- генеративные модели (generative models);

Стоит заметить, что *Deep Learning* привнес новый взгляд на проблему решения вопроса обучения больших полносвязанных графических моделей. В своих работах Hinton применяет современные методы цепей Монте-Карло, направляя сэмплированные частицы (*particles*), по направлению градиента поверхности правдоподобия модели. К сожалению его метод обучения ограничен только определенным классом моделей.

Как итог небольшого экскурса в историю ИНС, невооруженным взглядом можно заметить, что общий тренд их развития стремительно идёт по направлению к биоподобным моделям.

5 Новости с фронтов Neuroscience

Начиная с создания динамической модели аксона кальмара в работе Hodgkin и Huxley [10], в 50-ых сформировалась научная область изучающая поведение динамических систем и приведение их в соответствие с наблюдаемыми данными - *Computational Neuroscience*. Развитие компьютеров позволяют всё больше и больше понять тонкости взаимосвязи химических реакций в мозгу.

Neural Coding. До 90-ых в сообществе учёных была твёрдая уверенность, что информация в нейронах кодируется при помощи спайковой интенсивности (*rate coding*). В работе D. Warland [11] измерили временные статистики спайковой активности сверчка, на усы которого воздействовали слабым потоком воздуха, выяснилось, что нейронная активность наиболее всего коррелирует с временем между спайками. Также, работе J. Gautrais [12] показал, что потенциальное количество информации закодированной нейросетью растёт экспоненциально с ростом количества нейронов, если информация кодируется при помощи временной разницы между спайками.

Dendritic Computation. Роль дендритов много времени не была ясна и им не уделяли много внимания: многие симуляции организовывали во все без дендритов - синаптические связи подключали напрямую к клетке. В 2003 году появилась работа [13], где показывается нелинейность поведения дендритов и феномена появления, так называемого, кластера синапсов. Здесь [14] рассказывается, что используя только сеть дендритов и один нейрон была создана сеть распознающую лица.

Synaptic Plasticity. Пластичность синапсов, как ключ к обучаемости и интеллекту всегда был интересен ученым в Neuroscience и не только. Начиная с 1998 года появилась, уже классическая, модель пластичности синаптических связей [15] *Spike Timing Dependence Plasticity (STDP)*. Она во многом солидарна со правилом Хэбба [19], предсказанным им ещё в 1949 году – связи между нейронами, которые активны большую часть времени, усиливаются, однако правило не учитывает момент ослабления синаптических связей. Современное правило пластичности синапсов, биологически верное, построено на разнице во времени спайков pre-нейрона и post-нейрона. Если pre-нейрон произвел спайк, а потом произвел спайк post-нейрон, то синаптическая связь между pre-нейроном и post-нейроном усиливается. Если же post-нейрон произвел спайк раньше, чем pre-нейрон, то связь ослабляется. Модель, наиболее натурально описывающая этот феномен – экспоненциальная.

Существует множество работ [16][17][18], доказывающих, что рекуррентные сети, обучающиеся на таком правиле, способны к самоорганизации и кластеризации.

Polychronious Groups. Как гипотеза в работе Eugene Izhikevich [20], была предложена идея работы мозга на основе полихронов – особых групп нейронов, которые имеют место формироваться в больших симуляциях нейронных сетей (около 10^5 нейронов). Полихронные группы формируются в больших количествах и в разных качествах, есть группы временные, которые умирают сразу, есть постоянные, но также есть определенные виды групп которые соответствуют определенному входному паттерну [21].

6 Основная гипотеза

Вместо того, чтобы порождать огромное множество различных алгоритмов машинного обучения для каждого конкретного случая зависимости факторов (секция 2), предполагается создать на базе *Reservoir Computing* (секция 3) один специальный сложный алгоритм, который обучается на корреляционных атрибутах нейронной активности.

Гипотеза подразумевает существование единого принципа работы когнитивных функций мозга и возможности в большей или меньшей степени приблизить принцип работы резервуара с динамической нейронной сетью к этому единому природному принципу, что как следствие, предполагает наличия у резервуара свойств обнаружения сложных зависимостей в данных, которые кодируются в том или ином виде.

7 Задача и модель для проверки гипотезы

В качестве задачи предлагается задача обработки большого корпуса текстов. Результат обработки будет в выделении наиболее вероятной тематики текста (*Topic modeling*). Этой задаче присуще сложные внутренние зависимости среди скрытых переменных, которые образуют смысл текста.

Чтобы решить данную задачу необходимо решить отдельные подзадачи

относительно каждой части модели:

Входные элементы:

- Найти наиболее оптимальный способ представления текста во входных элементах. Предлагается начать со схемы, где каждый входной элемент ответственен за определенный слог [22];
- Найти оптимальный размер порции текста и наиболее оптимальное время «чтения» резервуара этой порции;

Резервуар:

- Создать наиболее оптимальную структуру резервуара, с т.з. обучения и хранения информации. Предлагается осуществлять это при помощи методов из Теории Информации [23];
- Выбрать наиболее оптимальную структуру резервуара. Предлагается начать с модели работ Izhikevich [20] и постепенно усложнять её различными вариациями, частично описанных в секции 5;

Выходные элементы, метод обучения. Предлагается использовать два подхода к обучению, которые сочетают различные аспекты решения задачи:

- Обучение с учителем. Для обучения предлагается использовать структурный метод опорных векторов, в качестве дополнительной ограничивающей структуры ввести кроссзависимости на полихронные группы нейронов (*Polychronius groups*);
- Обучение без учителя. Предлагается использовать модель Айзинга (*Ising models*) из физики, которая успешно применяется в раскодировании нейронной активности [24][25], и обучать её вариационным методом из *Deep Learning – Contrastive Divergence*[7];

8 Заключение

Автор рассмотрел последние наработки в областях использования рекуррентных сетей, последние, наиболее близкие к биологическим характеристикам, модели *Computational Neuroscience* и выработал естественное продолжение в сторону усложнения моделей *Reservoir Computing* с целью создать перспективный подход в обработке больших данных с внутренними кросс-зависимостями.

Также стоит отметить, что дополнительным итогом исследовательской работы может стать новое понимание работы когнитивных функций мозга.

Список литературы

- [1] D. Koller and N. Friedman, Probabilistic Graphical Models: Principles and Techniques, 2009, MIT Press
- [2] Max-Margin Markov Networks, B. Taskar, C. Guestrin and D. Koller. Neural Information Processing Systems Conference (NIPS03), Vancouver, Canada, December 2003
- [3] Herbert Jaeger, Harald Haas; Harnessing Nonlinearity: Predicting Chaotic Systems and Saving Energy in Wireless Communication; Science 2 April 2004: Vol. 304 no. 5667 pp. 78-80
- [4] Maass, Wolfgang; Natschläger, Thomas; Markram, Henry (November 2002), "Real-time computing without stable states: a new framework for neural computation based on perturbations Neural Comput 14 (11): 2531–60
- [5] Sepp Hochreiter; The Vanishing Gradient Problem during Learning; INTERNATIONAL JOURNAL OF UNCERTAINTY, FUZZINESS AND KNOWLEDGE-BASED SYSTEMS
- [6] <http://minds.jacobs-university.de/mantas/code>
- [7] Hinton, G. E., Osindero, S. and Teh, Y. (2006) A fast learning algorithm for deep belief nets. Neural Computation, 18, pp 1527-1554
- [8] Christoph Adami, Charles Ofria, and Travis C. Collier, Evolution of biological complexity (1997), PNAS 2000 97 (9) 4463-4468; doi:10.1073/pnas.97.9.4463
- [9] Yoshua Bengio, Aaron Courville, Pascal Vincent (2012), Representation Learning: A Review and New Perspectives, CoRR abs/1206.5538
- [10] A. L. Hodgkin. A. F. Huxley. Chance and design in electrophysiology: an informal account of certain experiments on nerve carried out between 1934 and 1952. J Physiol, 263(1):1-21, Dec 1976.
- [11] David Warland, Michael Landolfa, John P. Miller, William Bialek; Reading Between the Spikes in the Cerebellar Filiform Hair Receptors of the Cricket; Analysis and Modeling of Neural Systems; 1992, pp 327-333
- [12] Jacques Gautrais, Simon Thorpe; Rate coding versus temporal order coding: a theoretical approach; 1998; Centre de Recherche Cerveau et Cognition UMR 5549, 133 route de Narbonne, 31062 Toulouse, France
- [13] Michael Hausser and Bartlett Mel.; Dendrites: bug or feature? Current Opinion in Neurobiology 2003, 13:372–383
- [14] <https://class.coursera.org/bluebrain-001/class/index>
- [15] Guo-qiang Bi and Mu-ming Poo; Synaptic Modifications in Cultured Hippocampal Neurons: Dependence on Spike Timing, Synaptic Strength, and Postsynaptic Cell Type; The Journal of Neuroscience, 15 December 1998, 18(24): 10464-10472;

- [16] Xiaoli Tao, Howard E. Michel. Data Clustering Via Spiking Neural Networks through Spike Timing-Dependent Plasticity. In Hamid R. Arabnia, editor, Proceedings of the International Conference on Artificial Intelligence, IC-AI 04, June 21-24, 2004, Las Vegas, Nevada, USA, Volume 1. pages 168-173, CSREA Press, 2004.
- [17] Belatreche, Ammar and Paul, Rakesh (2012) Dynamic cluster formation using populations of spiking neurons. In: The 2012 IEEE International Joint Conference on Neural Networks (IJCNN), , Brisbane, Australia. IEEE. 6 pp.
- [18] Bernhard Nessler, Michael Pfeiffer and Wolfgang Maass; STDP enables spiking neurons to detect hidden causes of their inputs; Computational, Information-Theoretic Learning with Statistics, Learning/Statistics & Optimisation, Theory & Algorithms; 2009
- [19] Hebb, D. O. (1949). The Organization of Behavior: A Neuropsychological Theory. ISBN 978-0805843002.
- [20] Eugene M. Izhikevich, Joe A. Gally, Gerald M. Edelman; Spike-Timing Dynamics of Neuronal Groups; (2004) Cerebral Cortex, 14:933-944
- [21] Joseph Chrol-Cannon, Andre Gruning and Yaochu Jin; The Emergence of Polychronous Groups under Varying Input Patterns, Plasticity Rules and Network Connectivities; IJCNN The 2012 International Joint Conference on Neural Networks, 2012-06-10 - 2012-06-15, Brisbane, QLD
- [22] Ilya Sutskever, James Martens, Geoffrey Hinton; Generating Text with Recurrent Neural Networks; ICML 2011.
- [23] Borst A, Theunissen FE.; Information theory and neural coding; Nat Neurosci. 1999 Nov;2(11):947-57.
- [24] Elad Schneidman, Michael J. Berry, Ronen Segev & William Bialek; Weak pairwise correlations imply strongly correlated network states in a neural population; Vol 440|20 April 2006 doi:10.1038/nature04701
- [25] Gasper Tkacik, Elad Schneidman, Michael J. Berry and William Bialeka; Ising models for networks of real neurons; arXiv:q-bio.NC/0611072 v1 22 Nov 2006