1986 г. Октябрь *Том 150, вып. 2*

УСПЕХИ ФИЗИЧЕСКИХ НАУК

СОВЕЩАНИЯ И КОНФЕРЕНЦИИ

53(048)

НАУЧНАЯ СЕССИЯ ОТДЕЛЕНИЯ ОБЩЕЙ ФИЗИКИ И АСТРОНОМИИ И ОТДЕЛЕНИЯ ЯДЕРНОЙ ФИЗИКИ АКАДЕМИИ НАУК СССР

(26—27 февраля 1986 г.)

26 и 27 февраля 1986 г. в Институте физических проблем им. С. И. Вавилова АН СССР состоялась совместная научная сессия Отделения общей физики и астрономии и Отделения ядерной физики АН СССР. На сессии были заслушаны доклады:

26 февраля

1. В. Л. Афанасьев. Связь структуры галактик с активностью их ядер.
2. В. Ф. Шварцман. Исследования по релятивистской астрофизике и космологии на 6-метровом телескопе.
3. И. М. Копылов. Спектральные наблюдения на 6-метровом телескопе двойных систем с релятивистскими компаньонами.
4. Л. И. Снежко. 6-метровый телескоп БТА: состояние и перспективы.

27 февраля

5. П. Г. Костюк. Работа нервной клетки.
6. В. Л. Дунин-Барковский. Многонейронные структуры: теория и эксперимент.
7. Л. Б. Иоффе, М. В. Фейгельман. Спиновые стекла и модели памяти.

Краткое содержание пяти докладов приводится ниже.

[538.9+155.3](048)

Л. Б. Иоффе, М. В. Фейгельман. Спиновые стекла и модели памяти. В последнее время обнаружились глубокие аналогии между проблемами, возникающими при исследовании моделей ассоциативной памяти, и задачами статистической механики неупорядоченных систем типа спиновых стекол. Происхождение этих аналогий можно понять, рассмотрев основные требования, выдвигаемые при описании ассоциативной памяти:

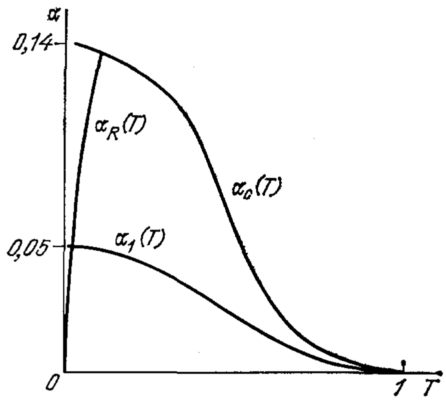
1) Система должна быть построена из большого числа N более или менее однородных элементов—*«нейронов»*, взаимосвязанных с помощью *«синапсов»*. В простейших моделях нейроны считаются двоичными элементами, а состояние каждого из них в данный момент времени зависит от состояний остальных за некоторый предшествующий интервал времени и от величин синаптических связей между ними.

2) Система должна быть способна к *классификаци*, т. е. 2^N входным сигналам (начальным состояниям системы) должно соответствовать существенно меньшее множество выходных сигналов, отвечающих стационарным состояниям («аттракторам»). Набор аттракторов задает информацию, записанную в системе памяти; переходный процесс от «входного» состояния к стационарному рассматривается как процесс «распознавания» полного «образа» по его малой части, заданной входным сигналом.

3) Система памяти должна допускать *последовательное обучение*, т. е. возможность добавления новых аттракторов (образов) без существенного искажения уже имеющихся.

4) Свойства системы должны быть *устойчивы* по отношению к случайным сбоям в работе отдельных «нейронов» и «синапсов».

Заметим теперь, что в физике хорошо известна система, удовлетворяющая всем указанным свойствам, кроме 3) : это — ферромагнитная модель Изинга, где роль нейронов играют изинговские «спины» $\sigma_i = \pm 1$, а величинам синаптических связей следует сопоставить энергии взаимодействия («обменные интегралы») J_{ij}. Динамика системы задается асинхронным процессом Монте-Карло релаксации энергии. Стационарные состояния такой системы подчиняются распределению Гиббса, что позволяет использовать хорошо разработанный аппарат статистической механики. При этом удобно рассматривать поведение системы в зависимости от температуры (соответствующей, для модели памяти, внешнему шуму в синаптических связях). При низких температурах система обладает всего двумя стационарными состояниями: все $\sigma_i = 1$ либо все $\sigma_i = -1$. Любое из начальных 2^N состояний быстро сходится к одному из этих двух. При слабом разбавлении (уничтожении части связей J_{ij}) свойства системы практически не меняются. Таким образом, для построения простейшей модели памяти необходимо иметь статистическую систему, аналогичную модели Изинга, но обладающую большим (при большом N) числом стационарных состояний. Таким свойством обладают широко исследующиеся в последние годы магнитные системы со случайными знакопеременными обменными интегралами J_{ij} — спиновые стекла [1]. Однако в классических моделях спиновых стекол число стационарных состояний слишком велико ($\sim \exp(\mathrm{const}\cdot N)$), а их конкретный вид неконтролируемым образом связан с видом заданной матрицы J_{ij}, поэтому стационарные состояния спинового стекла невозможно использовать для записи информации. Проблема решается в предложенной недавно Хопфилдом [2] модели с матрицей взаимодействия вида



$$J_{ij} = \frac{1}{N}\sum_{s=1}^{k} m_i^{(s)} m_j^{(s)},$$

$m_i^{(s)} = \pm 1$, причем N-вектора $m_i^{(s)}, m_i^{(s')}$ с $s \neq s'$ не коррелируют (такой выбор J_{ij} соответствует гипотезе Хебба о модификации величин синаптических связей в процессе обучения). При $k = 1$ эта модель сводится к модели Изинга заменой переменных $\sigma_i = \tilde{\sigma}_i m_i^{(1)}$; при температурах $T < T_c = 1$ имеются два стационарных состояния: $\sigma_i = \pm m_i^{(1)}$. При $1 < k < \frac{N}{4 \ln N}$ основные аттракторы совпадают с записанными «образами» $m_i^{(s)}$ [3], однако возникают дополнительные аттракторы, соответствующие смесям нескольких основных образов [4]. Области притяжения этих «фантомов» занимают малую часть фазового объема системы, так что все сформулированные выше требования к системе памяти оказываются выполненными. Для изучения пределов информационной емкости следует рассмотреть поведение системы при конечном значении $\alpha = \frac{K}{N}$ $(N \to \infty)$ [5,6]. В этом случае необходимо учитывать «интерференцию» различных образов $m_i^{(s)}$, приводящую к их искажению и, при достаточно большом α, исчезновению. Оказывается, что при $\alpha < 0{,}14$ (и низких температурах) система обладает аттракторами, минимумами свободной энергии $\tilde{m}_i^{(s)}$, близкими к исходным образам $m_i^{(s)}$: при $\alpha = 0{,}13$ «плотность ошибок» $p = 1 - 1/N \sum_i \tilde{m}_i^{(s)} m_i^{(s)} < 3\cdot 10^{-2}$. Наряду с этим имеется очень много аттракторов, отвечающих состояниям типа спинового стекла, некоррелированным с $m_i^{(s)}$. В интервале $0{,}05 < \alpha < \alpha_c = 0{,}14$ глобальный минимум свободной энергии реализуется на состоянии спинового стекла, область притяжения которого занимает большую часть фазового объема, однако начальные состояния с положительной проекцией на какой-либо образ $m_i^{(s)}$ релаксируют к $\tilde{m}_i^{(s)}$. При $\alpha < \alpha_1 = 0{,}05$ состояния $\tilde{m}_i^{(s)}$ являются глобально устойчивыми. Та же картина сохраняется при конечных температурах (см. рисунок), при $\alpha < \alpha_c(T)$ возможно устойчивое «распознавание» образа по некоторой его части. Следует заметить, что при $T > T_R(\alpha) = (8\alpha/9\pi)^{1/2} \exp(-1/2\alpha)$ состояние $\tilde{m}_i^{(s)}$, соответствующее $m_i^{(s)}$, единственно, в то время как при $T < T_R(\alpha)$ имеется большой набор $\tilde{m}_i^{(s)}$, близких к $m_i^{(s)}$ и выбираемых случайно. Это значит, что небольшой уровень шума стабилизирует работу системы. Таким образом, простейшая модель ассоциативной памяти построена. Оказывается, что ее свойства весьма устойчивы к ряду модификаций, желательных для лучшего соответствия с реальными нейронными системами: случайному «разбавлению» матрицы J_{ij}, ограничению значений J_{ij} всего двумя $(J_{ij}=\sqrt{k}/N)\,\mathrm{sgn}\,(\sum_{s=1}^{k} m_i^{(s)} m_j^{(s)}))$[7]. Наиболее существенным модельным ограничением является симметрия матрицы J_{ij}, необходимая для применения методов равновесной статистической механики, но не соответствующая данным нейрофизиологии. Предварительные численные эксперименты [2] показывают, что введение асимметрии J_{ij} не приводит к кардинальным изменениям. До сих пор обсуждалась проблема записи *некоррелированных* образов. Значительный интерес представляет вопрос о возможности построения памяти с *иерархической организацией* образов. Модель Хопфилда может оказаться удобным «первичным элементом» для построения таких систем [8].

СПИСОК ЛИТЕРАТУРЫ

1. Edwards S. F., Anderson P. W.//J. Phys. Ser. F. 1975. V. 5. P. 965.
   Fischer K.//Phys. Stat. Sol. Ser. b. 1985. V. 130. P. 13.
2. Hopfield J. J.//Proc. Nat. Acad. Sci. USA. 1982. V. 79. P. 2554; 1984. V. 81. P. 3088.
   Hopfield J. J., Feinstein D. I., Palmer R. G.//Nature. 1983. V. 304. P. 158.
3. Weisbuch G., Fogelman-Souie//de J. de Phys. Lett. 1985. T. 46. P. L623.
4. Amit D. I., Gutfreund H., Sompolinsky H.//Phys. Rev. Ser. A. 1985. V. 32. P. 1007.
   Веденов А. А., Левченко Е. Б.//Письма ЖЭТФ. 1985. Т. 41. С. 328.
5. Amit D. I., Gutfreund H., Sompolinsky H.//Phys. Rev. Lett. 1985. V. 55. P. 1530.
6. Feigel'man M. V., Ioffe L. B.//Europhys. Lett. 1986. V. 1.
7. Sompolinsky H. Preprint (to be published).
8. Dotsenko Vik. S.//J. Phys. Ser. C. 1985. V. 18. P. L1017.