

Архитектура информации

Одним из самых распространенных терминов не только в науке, но и в социальной сфере, и даже в быту, является "информация". Как известно, широкое распространение и признание в научных кругах этот термин получил благодаря работе Шеннона "Математическая теория связи" (1948). Но каждая наука и даже каждый человек субъективно трактуют этот термин, зачастую имея лишь его интуитивное понимание. Мы привыкли связывать с информацией некую абстрактную сущность, необходимую для коммуникации и познания мира. Без преувеличения можно сказать, что информация наряду с энергией являются одними из самых загадочных концепций, которые затрагивают множество научных дисциплин: физику, химию, биологию, философию, кибернетику, компьютерные науки, социальные дисциплины и даже теологию. Но если понятие энергии имеет формальные определения в физике и химии, то формального определения информации, которое было бы универсальным и определяло ее природу, мы не встретим. Очевидно, что сложность формализации этого понятия заключается именно в его междисциплинарности и многосторонности. Интерес к формальному определению информации усиливается в связи с возникшими проблемами обучения больших языковых моделей ИИ (LLM) и уменьшением уровня доверия к сгенерированной ими информации [1].

Но что мы знаем об информации с формальной, научной точки зрения? Теория информации Шеннона формализует понятие информации через энтропию: информация - разность энтропии до и после получения сообщения (т.е. информация — это мера уменьшения энтропии). Но его подход ограничен обработкой данных и не объясняет природу информации, т.е. ее "сущность" за пределами систем передачи данных. Алгоритмическая интерпретация информации Колмогоровым рассматривает ее как сложность объекта, которая может быть измерена с помощью алгоритмического описания – длины кратчайшей программы,

которая его генерирует. Такое рассмотрение информации не охватывает множество аспектов ее взаимодействия и потенциал измерений, сложно интерпретируется в физических, биологических, экономических и социальных системах.

Однако, определение информации в квантовой механике, которое дал фон-Нейман, уже связано с описанием состояния системы через энтропию, которая является аналогом энтропии Шеннона [2]. Существует также концепция Стивена Хокинга, который рассматривает понятие информации в контексте квантовой механики и черных дыр. В его работах информация понимается как характеристика квантового состояния системы и включает в себя *параметры, необходимые для полного описания состояния материи или излучения* [3].

1. Энтропия

Таким образом, можно сделать вывод, что в квантовой физике информация может интерпретироваться как *структура измеряемых параметров системы, которые описывают ее состояния или излучение энергии*. Например, уравнение Шрёдингера описывает состояния атома с помощью волновых функций (или функций состояния), которые зависят от координат всех его частиц [4].

Это позволяет сблизить понимание энтропии в термодинамике и квантовой механике с понятием информационной энтропии Шеннона. Давайте посмотрим на формулу термодинамической энтропии Больцмана и отметим некоторые известные факты

$$S = k_B \ln \Omega \quad (1),$$

где k_B – постоянная Больцмана, а Ω - количество возможных микросостояний, задаваемых *измеряемыми параметрами*: положением в пространстве, импульсом и энергетическим состоянием каждой частицы, при определенных

макроскопических условиях (состояниях) – измеряемых параметрах всей системы: энергии, объёме, давлении, температуре, или по-другому – это термодинамическая вероятность, которая устанавливает взаимосвязь между микроскопическими состояниями системы и ее макроскопическими свойствами. Сравним эту формулу с формулой информационной энтропии Шеннона (в этой формуле приведен натуральный логарифм вместо логарифма по основанию 2, который используется в оригинальной формуле Шеннона, для приведения к общему виду с формулой термодинамической энтропии)

$$H = - \sum_{i=1}^n p(x_i) \ln p(x_i) \quad (2),$$

где $p(x_i)$ – вероятность появления события x_i в некотором случайном процессе (например, последовательном процессе приема информации из некоторого источника). Мы можем увидеть, что для идеального газа термодинамическая вероятность рассматривается как отношение числа микросостояний при определенном распределении молекул к максимальному числу микросостояний при равномерном распределении молекул при заданных макроскопических условиях (т.е. определен порог максимальной энтропии). Эта термодинамическая вероятность может интерпретироваться как вероятность события, по аналогии с частотой появления символа в сообщении и распределении вероятностей всех символов в сообщении, в формуле Шеннона. Таким образом, термодинамическая энтропия S может рассматриваться как масштабированная (через постоянную Больцмана k_B) информационная энтропия H Шеннона.

В общем случае, при неравновероятных микросостояниях ($p_i \neq 1/\Omega$):

$$S = -k_B \sum_i p_i \ln p_i \quad (3).$$

Если же вероятности всех микросостояний p_i равны ($p_i=1/\Omega$), то энтропия Шеннона будет полностью совпадать с термодинамической энтропией

$$H \propto S \quad (4).$$

Таким образом, можно предположить, что энтропия Шеннона и энтропия Больцмана — это разные взгляды на одно и то же фундаментальное явление.

Аналогичные преобразования мы можем осуществить и с энтропией фон-Неймана. Для смешанных состояний эта энтропия описывается следующим выражением

$$S(\rho) = -\text{Tr}(\rho \log \rho) \quad (5),$$

где ρ — это матрица плотности, описывающая состояние системы, а Tr означает след матрицы, что отражает неопределённость в распределении вероятностей квантовых состояний

$$S = - \sum_i p_i \ln p_i \quad (6),$$

где p_i — собственные значения матрицы плотности ρ .

Таким образом, термодинамическую энтропию и квантовую энтропию можно было бы рассматривать как частные случаи более общего понятия информационной энтропии Шеннона, с дополнительными физическими интерпретациями. Взаимосвязь термодинамической и информационной энтропии является предметом современных исследований и научных дискуссий [5].

Однако влияние информации на уменьшение энтропии – *это одно из свойств информации, а не ее сущность*. И это свойство явно указывает на *энергетическую природу информации*. Понимание этой связи еще больше усилится, если мы вспомним принцип Ландауэра [6]: стирание информации в вычислительной системе связано с физическими затратами энергии, приводящими к выделению тепла в окружающую среду

$$E = k_B T \ln 2 \quad (7)$$

где k_B — постоянная Больцмана, T — температура окружающей среды (в кельвинах), $\ln 2$ — логарифм числа состояний при стирании одного бита. Таким образом, в соответствии с этим интуитивно понятным принципом, стирание информации увеличивает энтропию внешней среды, компенсируя уменьшение энтропии в системе. И, хотя принцип Ландауэра имеет лишь концептуальный характер и его физическое проявление связано со множеством ограничений, но он может восприниматься как косвенное доказательство общей природы энергии и информации.

Но, вернемся к информационной энтропии Шеннона. В этой формуле энтропия рассматривается как среднее количество информации в битах, которое необходимо для описания (представления с помощью бинарной системы счисления) случайного события или последовательности случайных событий. Чем менее вероятное событие, тем больше его энтропия и тем больше информации (количество информации определяется в битах или других структурных единицах информационной последовательности) оно несет, а вернее требует для снятия неопределенности. Следовательно, это событие имеет большую информативную ценность и для его представления в системе необходимо большее количество условных битов.

Мы можем сделать важный шаг в наших рассуждениях. Если мы, следуя Ландауэру, представим значение информационного бита как измеряемое значение энергии (или измеряемый параметр энергии), то вероятность этого значения в формуле Шеннона приобретает совсем другой смысл: эта вероятность показывает взаимосвязь или *взаимозависимость измеряемых параметров энергии в пространственно-временной структуре принимаемых (воспринимаемых) системой сигналов*. А если учесть, что в квантовой физике и химии взаимосвязь между структурными элементами энергии тоже представляет собой энергию – энергию взаимосвязи, то, следовательно, вероятность события в формуле Шеннона, *можно интерпретировать как уровень нашего (или системы) незнания о значениях параметров этой энергии взаимосвязи*. Переход от вероятностной взаимосвязи последовательности событий при определении информационной энтропии к энергетической взаимосвязи последовательности измеряемых параметров элементов энергии можно показать следующим образом

$$(x_1 p_{12} x_2 \dots x_{n-1} p_{n-1,n} x_n) \leftrightarrow (e_1 \gamma_{12} e_2 \dots e_{n-1} \gamma_{n-1,n} e_n) \quad (8)$$

где x_i – событие или бит информации в двоичной системе счисления, p_{ij} – условная вероятность двух последовательных событий в структуре принимаемого сообщения, e_i – измеряемый параметр элемента энергии (для формулы Шеннона может принимать значение 0 или 1), γ_{ij} – вероятность взаимосвязи двух последовательных элементов энергии, которая отображает уровень внешнего (по отношению к системе излучающей энергию) понимания этой взаимосвязи или степени уверенности в ее существовании. Эту вероятность γ_{ij} мы условно можем воспринимать как некий абстрактный эквивалент энергии взаимосвязи структурных элементов энергии. Это удобно делать с той точки зрения, что введение такого эквивалента позволяет воспринимать его как отображение внутренней энергии рассматриваемой системы, что, в свою очередь, позволяет

создавать модели изменения этой энергии. Аналогию такого подхода можно увидеть в генеративных искусственных нейронных сетях, где γ_{ij} могут интерпретироваться как весовые коэффициенты взаимосвязи токенов. Однако заметим, что такая вероятностная абстракция отражает лишь поверхностную взаимосвязь измеряемых параметров энергии, а точнее – их статистическую взаимосвязь. Корректность использования в этой формуле вероятности γ_{ij} для описания энергии взаимосвязи следует из аналогичного вероятностного представления этой энергии в методе Хартри-Фока в квантовой химии, а также при описании волновой функции, квантовых операторов и их собственных значений, туннельного эффекта и принципа неопределенности Гейзенберга в квантовой физике [7, 8]. Кроме того, такой подход также используется в молекулярной динамике и статистической механике. Такая интерпретация вероятности в формуле Шеннона также может быть основой для сближения его информационной энтропии с алгоритмической энтропией - Колмогоровской сложностью [9].

Обратим внимание еще на один важный момент. В предложенной интерпретации формулы Шеннона, x_i следует понимать не как измеряемый параметр элемента энергии, а как *обобщенное представление значения энергии в некоторой знаковой системе, т.е. абстракцию*. Причем, учитывая то, что одни и те же или различные виды и значения энергии мы можем представлять в различных знаковых системах, например, в бинарной системе – это значения битов, а в десятичной – это могут быть значения вероятностей, то это представление носит *субъективный характер*. Таким образом, в абстрактном представлении формулы информационной энтропии, x_i , p_{ij} или e_i , γ_{ij} приобретают смысл элементов алфавита информационной системы, которые мы или система можем наделять различным смыслом, но которые не имеют внутренней структуры, которую мы могли бы ассоциировать со структурой энергии.

Подведем некоторые итоги наших рассуждений относительно формулы информационной энтропии Шеннона:

1. Информационные элементы в формуле представлены в виде *абстрактных элементов алфавита* системы передачи и обработки информации (например, в битах или словах), а не в виде измеряемых параметров энергии системы (микро- или макросостояний), которая ее сформировала. Эта абстракция не позволяет рассматривать *внутреннюю структуру или сущность этих информационных элементов*, что является причиной отсутствия их контекстной (глубинной) взаимосвязи. Например, измеряемые параметры энергии взаимосвязаны, и эта взаимосвязанность определяется структурой внутренней энергии системы. Это хорошо видно в квантовой физике и химии.

2. Информационные элементы образуют *линейную пространственную-временную структуру*. Это отражает процесс обработки и передачи информации в информационных системах, однако существенно упрощает *структуру реальных взаимосвязей* между информационными или энергетическими элементами.

3. Взаимосвязь информационных элементов в формуле рассматривается только на *внешнем, статистическом уровне*, что не отражает *глубокую взаимосвязанность измеряемых параметров энергии (информации)* – зависимость макросостояний от микросостояний системы.

Эти ограничения являются причиной возникновения такого явления, как "*энтропийный разрыв*". Этот термин может обсуждаться в связи с возникающими проблемами генеративных ИНС [10]. Суть этого явления заключается не в "скачке" уменьшения информационной энтропии на выходе ИНС (это естественный процесс), а в том, что уменьшение информационной энтропии связано с увеличением вероятности взаимосвязи единиц информации, т.е. основано на статистической природе информационных взаимосвязей. Такой подход лежит в основе современной коннекционистской концепции построения и обучения ИНС. Однако, наиболее информативной является именно взаимосвязь с низкой

вероятностью, т.е. статистически редкая взаимосвязь. Это не позволяет современным ИНС генерировать новую или "креативную" информацию. Все, что создает (генерирует) ИНС, является статистически наиболее вероятным, т.е. зависимым от объема данных, используемых для обучения. Однако, с точки зрения новизны и креативности – это усредненные данные. Таким образом, редкие взаимосвязи, которые могли бы дать толчок к инновационным решениям или новым идеям, не получают должного веса, так как их вклад в общую вероятность распределения мал. Это и приводит к тому, что выход модели оказывается предсказуемым и ограниченным рамками усредненных статистических закономерностей. В этом и проявляется "энтропийный разрыв" между креативностью и посредственностью. Все это указывает на концептуальную ограниченность коннекционистского подхода в построении и обучении ИНС, т.к. любые попытки учета уменьшения статистической взаимосвязи в рамках энтропийного подхода неизбежно приведут к увеличению "хаоса". Это означает, что уровень доверия к такой сгенерированной информации будет падать. Некоторые исследователи ИИ говорят о том, что данные для обучения моделей ИИ уже закончились и связывают дальнейшее развитие моделей с использованием синтетических данных, сгенерированных самими моделями. Однако, информационная энтропия четко указывает на то, что такой подход, основанный на инбридинге данных, довольно быстро приведет к деградации информации и коллапсу модели, проявлению негативного эффекта аналогичного эффекту переобучения [11]. Кроме того, борьба с концептуальными ограничениями в рамках той же самой концепции может привести к некоторым улучшениям, но в конечном итоге неизбежно приведет к глобальному проигрышу. *Для эффективного решения возникших проблем нужна смена парадигмы, которая связана с определением информации и ее ролью в процессе построения системы.*

На основе приведенных рассуждений, мы могли бы сформулировать определение информации, объединяющее концепции термодинамической и

информационной энтропии, как *обобщенное субъективное представление энергии в определенной знаковой системе*. Однако, такое определение в большей степени является философским и не обладает достаточной степенью формализации для строгой научной концепции.

2. Энергетический ландшафт

Идея сближения в понимании общей природы энергии и информации может быть реализована в результате применения концепции *энергетического ландшафта* как к энергетическим (термодинамическим, диссипативным, квантовым и др.), так и к информационным, биологическим, экономическим или социальным системам. В физике энергетический ландшафт — это концепция, описывающая, как система эволюционирует в пространстве состояний, стремясь к состоянию с минимальной энергией. Например, в механике энергетический ландшафт может быть представлением потенциальной энергии системы как функции её пространственных координат, а в термодинамике он отражает свободную энергию в зависимости от параметров системы (температуры, давления, концентраций и т.д.). В таких ландшафтах, как правило, минимумы — локальные или глобальные точки ландшафта, соответствуют устойчивым состояниям системы, а максимумы и седловые точки представляют неустойчивые состояния системы. Различия между глобальными и локальными минимумами определяет сложность динамики или эволюции системы. Понятие энергетического ландшафта широко используется также в химии и биофизике (например, при фолдинге белков и катализе), статистической и квантовой механике. Безусловно, что реальные физические системы имеют сложный энергетический ландшафт, который описывает энергию системы в различных состояниях.

В общем случае, общая суммарная энергия системы U_{total} включает как локальные энергии отдельных i -х структурных элементов системы (юнитов) $U_{local,i}$, так и энергии $U_{bond,ij}$ взаимодействия между юнитами i и j .

$$U_{total} = \sum_i U_{local,i} + \sum_{i<j} U_{bond,ij}$$

(9).

Энергетический ландшафт описывает зависимость общей энергии U_{total} от конфигурации системы. Энергия взаимосвязей U_{bond} описывает взаимодействия между элементами системы и играет ключевую роль в формировании энергетического ландшафта. Примерами таких взаимосвязей могут служить: ковалентные или ионные связи в химических соединениях, силы межмолекулярного взаимодействия Ван-дер-Ваальса, энергия упругих взаимодействий в твёрдых телах (например, в кристаллах или молекулярных решётках), энергия магнитных или электрических диполей в системах с магнитными моментами или полярными молекулами. Эти взаимодействия часто имеют нелинейный характер, и их энергия зависит от расстояния между структурными точками.

На Рис.1 представлена общая структура энергетического ландшафта системы, которая описывается уравнением (9).

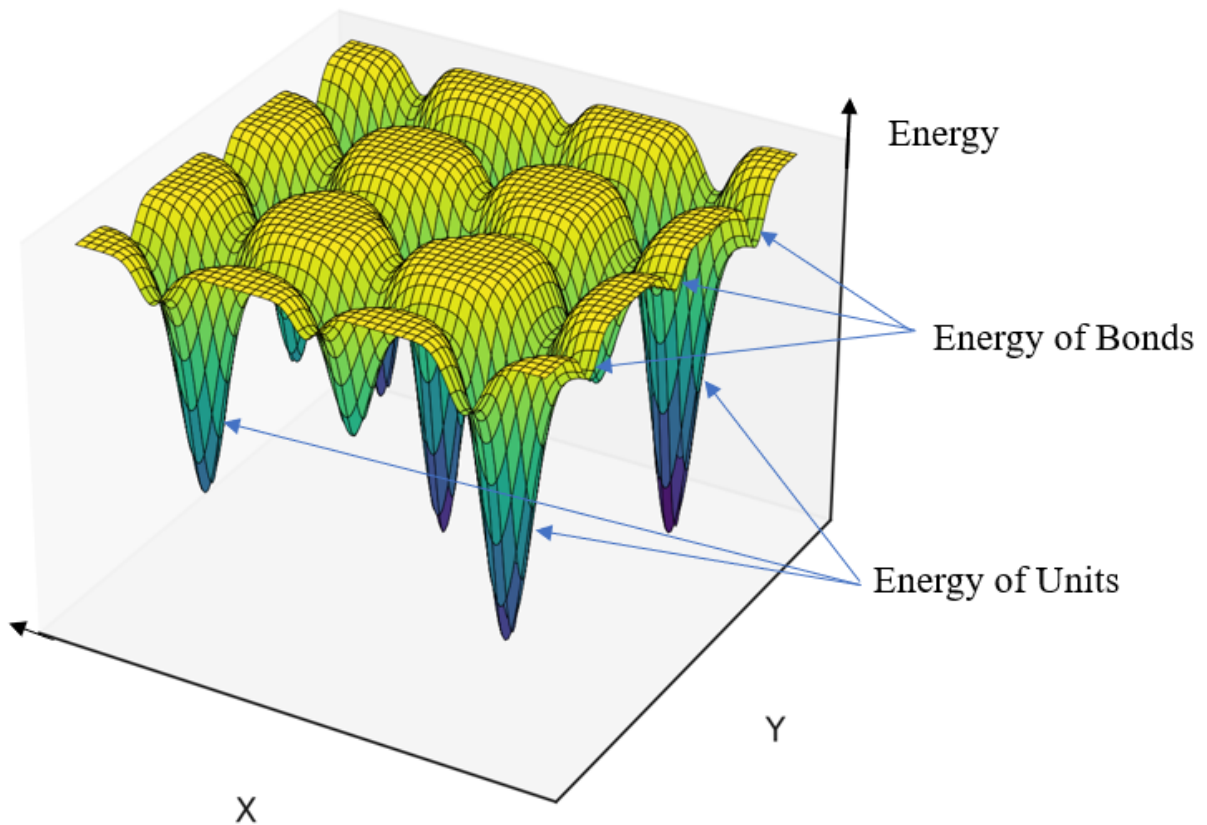


Рис.1. Общая структура энергетического ландшафта системы.

В общем случае под *стабильным внутренним энергетическим ландшафтом* El открытой энергетической системы будем понимать распределение внутренней энергии по структурным элементам (юнитам) U_i и связям между ними W_{ij} . без учета процессов необратимого рассеивания тепла ΔT , связанного с флуктуациями энергии юнитов вокруг их стабильного состояния, и без учета внешней энергии или потоков энергии между системой и окружающей средой.

$$El_{rest} = \min(\sum_{i=1}^N U_i + \sum_{i=1}^N \sum_{j=1}^N W_{ij}) = Tr1 \quad (10).$$

В данной формулировке, назовем этот энергетический ландшафт – *ландшафтом системы в состоянии покоя или энергетического равновесия* El_{rest} , а

общую энергию системы в этом состоянии – *порогом покоя* Tr_1 . В этом состоянии система достигает *максимальной энтропии* для данного макросостояния, при котором не наблюдается никаких потоков энергии, массы или вещества. Например, при термодинамическом равновесии все термодинамические силы, такие как градиенты температуры, давления или химического потенциала, равны нулю, а все процессы, которые могли происходить внутри системы (например, химические реакции, теплопередача), завершены. Даже если на микроскопическом уровне системы продолжается флуктуация энергии - изменение параметров внутренней энергии системы (например, движение частиц), то система все равно будет находиться в *состоянии равновесия (покоя)*, которое определяет максимум возможных микросостояний системы для данного макросостояния, которые не приводят к его изменению. Т.е. при стабилизации структуры и минимизации ее внутренней энергии мы говорим о максимальном уровне энтропии системы для определенного макросостояния. Состояние покоя или равновесия системы является идеализированным для закрытой термодинамической системы. Когда такая система достигает состояния покоя, то она имеет максимальную энтропию.

Если же мы говорим об открытой системе, в которую постоянно поступает внешняя энергия, что позволяет ей поддерживать неравновесное состояние и создавать упорядоченные структуры, то для такой системы, называемой диссипативной системой, наблюдается *локальное уменьшение ее внутренней энтропии в результате структуризации - возникновении диссипативных структур*. При этом общее макросостояние такой системы может оставаться неизменным, что также ведет к увеличению общей термодинамической энтропии в соответствии со вторым законом термодинамики. Эти процессы составляют основу теории диссипативных систем Пригожина [12].

Примерами систем в состояний покоя, в зависимости от интерпретации энергии, являются: газ в закрытом изолированном сосуде, который достиг термодинамического равновесия (температура, давление и плотность равномерны

по всему объему), молекула в вакууме, изолированный биологический нейрон в состоянии покоя, искусственная нейронная сеть в состоянии покоя (т.е. в отсутствии входных сигналов и изменения внутренних параметров). Можно также ввести аналогию экономической или социальной энергии покоя в виде денежных средств предприятия, хранимых в банке, системы культурных и моральных ценностей в обществе и др.

Однако заметим, что это состояние указывает на важный фундаментальный предел – *энергетический порог покоя системы, определяющий минимум внутренней энергии системы и максимум энтропии* для данной структуры и определенного множества макросостояний системы. Энергетический ландшафт системы в состоянии покоя зависит от внутренней физической структуры системы. Таким образом, этот ландшафт образует энергетический фундамент системы, который определяет основу ее устойчивости к внешним энергетическим воздействиям.

Минимум энергии ландшафта в состоянии покоя обеспечивается *минимизацией физической структуры системы и/или минимизацией внутренней (потенциальной) энергии юнитов и их взаимосвязей*. Минимизация структуры может быть связана с ее динамической адаптацией при изменении внешнего энергетического воздействия или с эволюционным изменением структуры, как это, например, может происходить в точках бифуркации диссипативных систем или в биологических системах. Однако, если мы говорим о минимизации внутренней энергии юнитов и их взаимосвязей, то мы говорим о том, что *эта энергия также имеет структуру*. Примеры описания такой структуры энергии мы можем найти в квантовой механике и квантовой химии – это уравнения Шрёдингера, метод Хартри-Фока, потенциал Леннарда-Джонса, теория функционала плотности (DFT) и др. При этом, исходя из теории квантовой механики, энергия взаимосвязей E_{bond} уменьшает суммарную энергию отдельных атомов E_{atoms} в общей структуре энергии молекулы $E_{molecule}$.

$$E_{bond} = \sum E_{atoms} - \sum E_{molecule} \quad (11).$$

Это говорит о том, что часть потенциальной энергии юнитов используется для создания взаимосвязей между ними. Таким образом, энергетический ландшафт системы в состоянии покоя состоит из "ям" потенциальной энергии, представленных юнитами и "каналов" для переходов потенциальной энергии между "ямами", представленных в виде взаимосвязей (Рис.1). Потенциальная энергия в "ямах" и "каналах" имеет структуру, состоящую из энергетических уровней или подуровней. Эти потенциалы практически формируют энергетическую емкость системы и указывают на уровень "дна" этой емкости относительно некоторой точки отсчета ("плато" ландшафта). Например, для атома потенциальная энергия электрона в поле ядра указывает на "дно" его энергетического состояния относительно энергии ядра. Мы образно можем представить этот уровень, как положение некоторой точки геодезического ландшафта над уровнем моря.

Таким образом, образование стабильных структур ведет к уменьшению внутренней энергии системы в состоянии покоя.

Если мы применим это описание энергетического ландшафта покоя к ИНС, то увидим, что:

1. Для большинства типов моделей, *энергетический ландшафт покоя* статичен и, как может показаться, состоит исключительно из весовых коэффициентов, которые мы можем интерпретировать как энергии взаимосвязей нейронов (юнитов), причем, эта энергия в моделях прямого распространения сигналов учитывается только в одном направлении. Однако это не так, потому что в ИНС взаимосвязи существуют "физически", но их не существует как отдельных энергетических элементов. Если мы посмотрим на модель МакКаллока-Питтса, которая лежит в основе большинства моделей нейронов ИНС,

$$Y = f\left(\sum_{i=1}^n w_i x_i + Q\right) \quad (12),$$

где Y – реакция нейрона, f – активационная функция, w_i – весовой коэффициент i -го входа нейрона, x_i – входной сигнал, подаваемый на i -й вход, Q – порог активации нейрона, то мы увидим, что весовые коэффициенты играют роль модуляторов входных сигналов (входной энергии) нейрона, а не самостоятельных энергетических элементов - взаимосвязей. Однако, если учитывать, что входные сигналы равны значениям битов (как в классической модели), то с энергетической точки зрения именно сумма весовых коэффициентов могли бы определять структуру энергии и возможную глубину потенциальной "ямы" юнита энергетического ландшафта ИНС и, следовательно, общий энергетический ландшафт системы в состоянии покоя (10). Т.е.

$$U_{jrest} = \sum_{i=1}^N w_{ij} \quad (13),$$

$$El_{rest} = \sum_{j=1}^N U_{jrest} = \sum_{i=1}^N \sum_{j=1}^N W_{ij} \quad (14),$$

где U_{jrest} – энергия j -го юнита (нейрона) в состоянии покоя.

Действительно, если вход $x_i = 0$, то эта входная энергия не покидает i -й подуровень потенциальной "ямы" юнита $w_{ij}x_i = 0$, а если $x_i = 1$, то энергия выхода с этого подуровня составляет w_{ij} , т.е. сумма весовых коэффициентов в этом случае определяет максимальную "глубину" потенциальной "ямы". В целом, проводя

аналогию с энергетической системой, например с атомом, мы могли бы сравнить работу нейрона с процессом ионизации атома, когда линейная комбинация входов $\sum w_i x_i + Q$ моделирует выход электрона из атома, т.е. образование свободной энергии, тогда как $w_i x_i$ моделирует переход одного электрона на другой энергетический подуровень, где w_i определяет энергию подуровня в состоянии покоя. Однако это лишь условная аналогия, которая, в данном случае, не имеет общей энергетической основы. Таким образом, модель формального нейрона – информационная модель, основанная на математической, а вернее статистической абстракции информационных процессов, и в классической интерпретации далекая от описания термодинамических процессов.

2. Значения отдельных весовых коэффициентов для большинства моделей ИНС могли бы интерпретироваться как условная внутренняя энергия системы, которая не имеет структуры и не связана с входной энергией системы. Действительно, каждый нейрон просто равномерно распространяет (дублирует) свою выходную энергию на все нейроны следующего слоя. Это распространение энергии можно сравнить с ее рассеиванием. Однако, в реальности такое "рассеивание" в ИНС (если представить ее физическую реализацию) связано с дополнительными энергетическими затратами для сохранения уровня энергии реакции нейрона на входах множества нейронов следующего слоя. Весовые коэффициенты входов связаны с другой энергией – псевдоэнергией. Однако, эта псевдоэнергия не управляет распределением внешней энергии по внутреннему энергетическому ландшафту системы, как могло бы казаться, а определяет согласованные потенциальные вклады отдельных элементов входной энергии нейронов, направленные на преодоление их порогов активации. Эта псевдоэнергия является энергией ошибки, которая минимизируется в процессе итерационного обучения. Таким образом весовой коэффициент формально определяет "правильный", с точки зрения минимизации энергии ошибки, вклад энергии входа

в преодоление порога активации, т.е. выхода энергии из "потенциальной ямы" для достижения требуемого значения реакции нейрона и ИНС в целом.

3. Входная энергия ИНС, нормализована и ограничена системой представления, которую условно можно сопоставить с сенсорным восприятием, сегментирующим и преобразующим внешнюю энергию во внутреннюю. Таким образом, входная и выходная энергии представлены еще одним видом псевдоэнергии, *выраженной в символах внутреннего алфавита системы*, например, в символах двоичной или десятичной системы счисления.

Такое представление входной энергии, наряду с другими указанными ограничениями, ведет к формированию детерминированного множества возможных *микросостояний системы*, которые для ИНС определяются множествами весовых коэффициентов в состоянии покоя, и распределением входной энергии по сети, т.е. реакциями нейронов, в состоянии обработки информации.

Тогда, множество *макросостояний* ИНС определяется реакциями нейронов выходного слоя и, с энергетической точки зрения, образует свободную энергию системы. В этом смысле результат работы сети (её выход) отражает новое стабильное состояние системы (в котором она минимизировала свою энергию), обусловленное взаимодействием входной энергией (входным вектором сигналов) с энергетическим ландшафтом.

Можно сказать, что ИНС в состоянии инференса (работы обученной сети) находится в *состоянии стабильного энергетического неравновесия*. В этом состоянии система выходит из состояния покоя, но структура ее внутреннего энергетического ландшафта покоя не нарушается. Система поддерживает устойчивую внутреннюю (физическую) структуру и динамику, а также заданные макросостояния.

Это состояние связано с существованием второго фундаментального порога – порога активации системы Tr_2 . Этот порог задает внутреннюю метрику системы,

позволяя определить глубину "ям" и "каналов" энергетического ландшафта системы. Величина локального порога активации юнита определяет *глубину его потенциальной энергетической "ямы"*, т.е. его энергетическую емкость и зависит от внутренней структуры его энергии. Например, на квантовом уровне электрону нужно преодолеть несколько энергетических подуровней, чтобы оторваться от атома, т.е. преодолеть барьер ионизации, который мы можем рассматривать как аналог порога активации. Для разных систем будут существовать разные пороги активации: для биологических нейронов - потенциал действия, для искусственного нейрона - активационная функция, для динамических систем порог активации может быть описан функцией или функционалом Ляпунова или, в зависимости от типа системы, сложности взаимодействий и свойств системы могут быть использованы другие методы, такие как: матрицы Якоби, фазовые диаграммы, энергетические барьеры, обратная связь и статистические методы, критические явления и др. Таким образом, тип и структура внутренней энергии задает правила или условия активации системы. В общем случае

$$E_{act(i)} = (U_{min(i)} + \Delta U_i) > Tr2_i \quad (15),$$

$$\Delta U_i = F(X) \quad (16),$$

где $E_{act(i)}$ – энергия активации i -го юнита, $U_{min(i)}$ – потенциальная энергия покоя i -го юнита, ΔU_i – дополнительная энергия i -го юнита, $Tr2_i$ – порог активации i -го юнита – предел его потенциальной емкости, $F(X)$ – функция распределения входной энергии по юнитам.

Представим себе поверхность с ямами, которые соединены ложбинами или каналами, и, возможно, холмами. Если мы выльем воду на эту поверхность, то увидим процесс распределения воды по поверхности с заполнением ям и каналов. Это представление наглядно отображает процесс распределения энергии по

энергетическому ландшафту системы. Если есть приток энергии в систему и нет ее эффективного отвода, то энергия переполняет энергетическую емкость системы и энергетические взаимосвязи между юнитами могут быть разрушены, т.к. их потенциальная емкость намного меньше емкости юнитов. Эта емкость взаимосвязей определяет третий фундаментальный порог структурной стабильности или устойчивости системы $Tr3$.

$$E_{str(ij)} = (W_{min(ij)} + \Delta U) \leq Tr3_{ij} \quad (17),$$

где $E_{str(ij)}$ – энергия стабильной взаимосвязи между юнитами i и j , $W_{min(ij)}$ – минимальная потенциальная энергия взаимосвязи между юнитами i и j , $Tr3_{ij}$ – порог структурной устойчивости данной взаимосвязи. Если разрушение взаимосвязей превысит порог структурной устойчивости всей системы, то это может определять точку бифуркации системы, т.е. точку изменения ее состояния или развития, что хорошо иллюстрируется в теории диссипативных систем Пригожина.

В ИНС порог активации представлен правилом или функцией формирования реакции искусственного нейрона. В большинстве современных моделей нейронов эта активационная функция (например, сигмоидная или ReLu) связана с нормализацией реакции и не отражает уровень превышения порога активации нейрона, но может служить аналогом уровня потенциальной энергии нейрона в состоянии активации.

Существование двух порогов $Tr1$ и $Tr2$ определяют устойчивость любой системы к флуктуациям в межпороговом пространстве. Превышение порога активации на определенную величину демонстрирует одну и ту же стабильную реакцию системы, например, на выходе системы. Для перехода системы в нестабильное состояние нужна внешняя энергия, меняющая энергию покоя, порог активации или преодолевающая порог структурной устойчивости. Так как

энергетический ландшафт состоит из множества компонентов, то флуктуация энергии, нарушающая локальные элементы порога $Tr3$, может не привести к разрушению всего ландшафта, но увеличение количества таких флуктуаций может привести к точке бифуркации.

Такой подход к рассмотрению изменений состояния энергетического ландшафта системы позволяет по-новому взглянуть на классификацию типов систем:

- если система может адаптироваться под эти изменяющиеся условия в результате изменения порогов или параметров внутренней энергии, то она является адаптивной;

- если адаптация происходит в результате перестройки своего энергетического ландшафта покоя, связанного с изменением внутренней структуры, то она является эволюционной;

- если система способна создавать новый энергетический ландшафт покоя на основе синтеза внутренней структуры, то она является самоорганизующейся.

Таким образом, порог активации определяет максимально допустимую флуктуацию внутренней энергии для сохранения состояния покоя (равновесия). При превышении этого предела, который определяется структурой внутренней энергии, происходит переход системы в стабильное неравновесное состояние. Порог бифуркации зависит от структуры самой системы, порогов покоя и активации. Если нарушается стабильность реакции системы при увеличении флуктуаций, это может свидетельствовать о приближении к порогу бифуркации. Далее система может перейти к адаптации, связанной с самоорганизацией, либо к состоянию деградации, т.е. неравновесному состоянию с нарушением внутренней энергетической и физической структуры.

Можно сделать важный вывод о том, что структура энергии является основой не только построения внутреннего энергетического ландшафта системы, но и процессов ее адаптации, эволюции и самоорганизации.

Если мы говорим о том, что адаптация системы может быть не только на основе обратной связи, но и на основе глубоких внутренних механизмов преобразования энергии, связанных с изменением порогов и энергетического ландшафта системы, то это может существенно расширить теорию систем и кибернетику. Кроме того, это может изменить представление об обучении искусственных нейронных сетей, открывая путь к новому взгляду на их самообучение.

3. Модель Хопфилда

Первой и самой известной моделью ИНС, которая использует концепцию энергетического ландшафта и функцию энергии, является модель Хопфилда [13, 14]. В этой модели была предпринята попытка сближения теории информации и физики. Проанализируем, насколько это сближение обосновано.

С точки зрения энергетических состояний модели, наиболее известные физические интерпретации концепции Хопфилда, утверждают следующее:

1. Энергетический ландшафт интерпретируется как множество возможных состояний нейронов.

2. Минимизация, а вернее уменьшение энергии при изменении состояния как в отдельном нейроне, так и в целом в сети иллюстрируется функцией энергии Хопфилда.

3. Локальный аттрактор, который определяет "правильное" состояние одного нейрона, как элемент глобального аттрактора, имеет "бассейн притяжения", который определяется состояниями нейрона, приближающимися к устойчивому значению, которое сопоставляется с запомненным элементом паттерна. Недостижение нейроном состояния локального аттрактора характеризуется как нежелательное явление - "ловушка" локального минимума энергии. Это явление может возникать из-за интерференции между запомненными паттернами.

4. Из этих утверждений делается вывод о существовании множества глобальных аттракторов с минимальной энергией, каждый из которых сопоставляется с запомненным паттерном.

В качестве прототипа модели Хопфилда используется модель Изинга для описания свойств ферромагнитных материалов [15]. Модель Изинга является физической моделью и ее основными характеристиками являются энергия, энтропия и температура, которые используются для описания поведения системы из большого числа взаимодействующих спинов в магнитных материалах. В этих материалах взаимодействия между магнитными моментами (спинами) приводят к сложным и неупорядоченным состояниям. Спины представляют собой ориентацию магнитных моментов атомов и связаны с энергией их взаимодействий. Эта модель была разработана для понимания того, как локальные взаимодействия между соседними спинами могут приводить к глобальному упорядочению в системах, например, к фазовому переходу ферромагнетика. Это упорядочение зависит от температуры: при низких температурах спины имеют тенденцию выстраиваться в одном направлении (упорядоченное состояние), а при высоких — они становятся случайно ориентированными (дезупорядоченное состояние). Данная модель не учитывает все виды энергии, ее диссипацию или образование свободной энергии, т.е. она не является полной термодинамической моделью.

Если эту модель рассмотреть более детально, то *она изучает коллективное поведение спинов S_i , которые могут принимать значение ± 1* . Взаимодействие между спинами описывается матрицей константных или случайных величин J_{ij} . Можно сказать, что J_{ij} — *это условная энергия взаимосвязи, которая может отображать корреляцию спинов: положительную, т.е. параллельную ориентацию спинов, или отрицательную - дезориентацию спинов*. Важно то, что это действительно энергия взаимосвязи, а не элемент внутренней энергии юнита (атома). Однако, в классической модели Изинга все J_{ij} имеют значение $+1$, т.е. аксиоматически задается, что состояния спинов стремятся к одному устойчивому

состоянию с параллельной ориентацией. Текущие состояния спинов зависят от температуры, влияющей на вероятность изменения их ориентации в соответствии с распределением Больцмана: чем ниже температура, тем выше вероятность параллельной ориентации.

В процессе динамического изменения состояний спинов формируется сложный энергетический ландшафт системы. Динамика модели Изинга описывается через энергетический функционал - Гамильтониан Изинга H .

$$H = -\frac{1}{2} \sum_{i \neq j} J_{ij} S_i S_j - \sum_i h_i S_i \quad (18),$$

где h_i — внешнее поле.

Энергия системы уменьшается по мере ее эволюции, пока не будет достигнут минимум, характеризующий устойчивое состояние системы с параллельной ориентацией спинов. При отсутствии внешнего поля $h_i = 0$, это состояние полностью определено матрицей весовых коэффициентов взаимосвязи спинов. Таким образом, в *стабильном состоянии покоя системы энергетический ландшафт, в соответствии с формулой (18), будет представлен матрицей весовых коэффициентов со значениями +1* и будет отображать состояние системы с минимальной энергией и максимальной термодинамической энтропией.

Гамильтониан Изинга является дискретным аналогом функционала Ляпунова, который используется для описания непрерывных устойчивых состояний динамических систем [16].

В модели Хопфилда используется идея модели Изинга, адаптированная для решения задачи построения ассоциативной памяти.

Функция энергии модели Хопфилда, которая уменьшается с каждым шагом обновления состояний нейронов, описывается следующим функционалом

$$E = -\frac{1}{2} \sum_{i \neq j} w_{ij} s_i s_j - \sum_i \theta_i s_i \quad (19),$$

где s_i – состояния нейронов (+1 или -1), w_{ij} – весовые коэффициенты связей между нейронами, θ_i – пороговые значения.

Одним из основных отличий этих моделей является то, что в классической модели Изинга изменения состояний спинов происходят случайным образом с использованием алгоритма Метрополиса (метод Монте-Карло) [17], который моделирует взаимосвязь внешней и внутренней энергии системы, в то время как в модели Хопфилда изменения состояния нейронов происходят детерминировано на основе сравнения входных сигналов со знаками взаимных корреляций, заданными в виде суммы весовых коэффициентов. Изменения состояний нейронов стремятся к восстановлению определённого состояния (паттерна) в результате рекуррентной архитектуры модели. Т.е. архитектура рекуррентных связей сети Хопфилда выполняет роль своеобразной "метлы" для поддержания ее динамики.

В модели Изинга энергетический порог "переворота" спина зависит от температуры T и изменения энергии ΔE . Этот порог определяется вероятностью P , с которой система "решит", стоит ли принять новое состояние, которое может иметь большую энергию.

$$P(\Delta E) \propto e^{-\Delta E/k_B T} \quad (20),$$

где k_B – постоянная Больцмана.

Вероятность P , в соответствии с распределением Больцмана, играет роль аналога энергетического порога, определяющего изменение состояния спина. Таким образом, взаимосвязь энергетического порога с термодинамическими параметрами определяет физическую основу модели Изинга. Мы видим, что в

модели Изинга этот порог *динамический*, тогда как в классической модели Хопфилда порог θ_i изменения состояния нейрона *статический* и равен 0. *Этот порог в модели Хопфилда, как и другие параметры его модели, никак не связан с термодинамикой, что определяет модель Хопфилда как информационную.*

Давайте обратим внимание на тот факт, что при описании функционирования модели Хопфилда используется понятие "энергетический ландшафт" системы как совокупность состояний нейронов, но не дается его определения для равновесного состояния покоя. Однако, если это система энергетическая, то в ней должен существовать ландшафт в состоянии покоя.

Легко увидеть, что в соответствии с функцией энергии Хопфилда, энергетический ландшафт системы в состоянии покоя образуют суммы весовых коэффициентов на входах нейронов, т.е. именно весовые коэффициенты приобретают смысл *внутренней энергии сети в состоянии покоя или равновесия*. Действительно, в соответствии с правилом обучения Хэбба, эти весовые коэффициенты формируются однократно в процессе обучения и далее динамически не обновляются

$$w_{ij} = \frac{1}{N} \sum_{k=1}^p x_i^k x_j^k \quad (21),$$

где x_i^k – значение реакции i -го нейрона в k -м образце, N – количество нейронов в сети, p – количество образцов для запоминания. Таким образом, в модели Хопфилда веса w_{ij} формируются на основе корреляции паттернов, которые сеть должна запомнить, а не на основе фундаментальных энергетических взаимодействий.

Тогда, в соответствии с формулой (19), минимальная энергия сети в состоянии равновесия E_{min} , т.е. теоретически, если бы все состояния нейронов (их

реакции) идеально коррелировали между собой на основании значения их весовых коэффициентов, будет равна

$$E_{min} = -\frac{1}{2} \sum_{i \neq j} w_{ij} \quad (22).$$

Таким образом, сеть имеет один *глобальный минимум энергии системы в состоянии покоя*, который является теоретическим пределом, и практически в работающей сети не достижим в силу различных типов корреляций для разных сигналов и особенности алгоритма формирования весовых коэффициентов. Можно сделать вывод, что действительно для каждого запомненного паттерна будет свое минимальное значение функции энергии, которое на практике будет зависеть от степени искажения запомненных образцов. Это можно интерпретировать как существование *глобальных энергетических минимумов для каждого запомненного образца*.

Это объясняется тем, что знаки сумм весовых коэффициентов нейрона указывают на "правильную" или "желательную" корреляцию входных сигналов с запомненными паттернами – положительную или отрицательную. Сумму взвешенных входных сигналов нейрона можно интерпретировать как "*корреляционное голосование*", в результате которого побеждает решение об изменении или сохранении текущего состояния нейрона. Результат этого "голосования", отражающий максимально возможную корреляцию сигналов, будет зависеть не только от значений весовых коэффициентов, но и от текущих состояний нейронов.

Если не будет достигнута полная корреляция в соответствии с весовыми коэффициентами, то система не достигнет желаемого глобального минимума, соответствующего запомненному паттерну, что и будет отражено в формуле энергии как нежелательное явление локального энергетического минимума. Таким

образом, в модели Хопфилда *функция энергии может увеличиваться на уровне отдельных нейронов, если нарушаются локальные корреляции*. Однако глобальная динамика сети, с учетом возможных ограничений, будет направлена на уменьшение общей функции энергии. Это подчеркивает, что модель Хопфилда работает с глобальной энергетической функцией, а локальные отклонения в отдельных элементах могут компенсироваться общим поведением системы.

Таким образом, можно сделать следующие выводы:

1. Модель Хопфилда, в отличие от модели Изинга является информационной, т.е. абстрагированной от какой-либо термодинамической или любой другой энергетической интерпретации. Поэтому, при ее рассмотрении было бы корректнее использовать термин "условная" или "псевдо" энергия. Однако использование понятий "энергетический ландшафт" и "функция энергии", тем не менее, являются мощными концептуальными шагами. *Минимум функции энергии Хопфилда демонстрирует достижение стабильного состояния и уменьшение информационной энтропии системы.*

2. Классическая модель Хопфилда имеет пространственный линейный энергетический ландшафт системы в состоянии покоя, представленный суммами весовых коэффициентов нейронов. Это объясняется физической структурой сети. В модели Изинга этот ландшафт отображает структуру решетки атомов. При рассмотрении модели Хопфилда возникает дилемма интерпретации энергии весовых коэффициентов: их можно было бы отнести к элементам внутренней энергии нейронов (формула (10)) или к энергии взаимосвязи нейронов (формула (14)), т.к. они являются функцией от эталонных значений их внутренних состояний в соответствии с формулой (21). Однако, в силу структуры модели МакКаллока-Питтса (12, 13, 14), их все же следует отнести к элементам внутренней энергии юнитов (нейронов). В этом случае коэффициент $\frac{1}{2}$ в формуле энергии Хопфилда (19) является лишним, в отличие от Гамильтониана Изинга (18),

где весовые коэффициенты действительно отображают энергию взаимосвязи атомов.

3. Функция энергии Хопфилда отделена от процесса управления динамикой сети и служит только иллюстрацией процесса. Поэтому метафора "аттракторов" информационных или энергетических не является корректной, т.к. в модели нет энергии или энергетического состояния, которое бы "притягивало" другие энергетические или информационные состояния. Т.е. образное представление работы сети как движение шара по поверхности энергетического ландшафта с попаданием в "яму" с минимальной энергией не соответствует реальным процессам.

Это утверждение обосновывается тем, что метафора "элемента глобального аттрактора" могла бы относиться только к сумме весовых коэффициентов каждого нейрона, как к единственной явной, а не виртуальной, константной величине. Но проблема в том, что в этом же качестве эта сумма выступает для всех "глобальных аттракторов" и все корреляции между "элементами глобального аттрактора" взаимосвязаны матрицей этих весовых коэффициентов. Решение о смене состояния нейрона зависит от значений входных сигналов, т.е. "аттрактор" должен быть динамическим и меняться в зависимости от входных сигналов. Нужно сказать, что в теории динамических систем, в нелинейной динамике существуют концепции, которые описывают такие явления, но для этого система должна иметь: мультистабильные состояния и механизм переключения между ними; или аттракторную метаморфозу с постепенным или внезапным изменением формы или типа аттрактора; или хаотическую эволюцию аттрактора и др., что возможно для сложных систем с изменяющимися параметрами или взаимодействиями. Тогда, более корректно рассматривать не метафору "аттрактора", а истинную корреляционную природу этих весовых коэффициентов и их суммы. Можно назвать этот процесс – *корреляционным голосованием*, когда большинство взвешенных входных сигналов "голосует" за положительную или отрицательную

корреляцию выходного сигнала, т.е. за принятие решения о переходе состояния в +1 или -1.

4. Если функция энергии Хопфилда (19) его модели указывает на уменьшение условной энергии, связанное с уменьшением информационной энтропии, то аналогичная функция энергии Изинга (18) также на это указывает. Таким образом, можно сделать вывод, что *физическая спин-стекольная система производит не только термодинамическую, но и информационную энтропию*. Это согласуется с теорией квантовой информационной энтропии и теорией диссипативных систем Пригожина.

В целом, функционирование модели Хопфилда можно было бы проще и корректнее описать без использования энергетических терминов: конечное стабильное состояние системы зависит от начального состояния (входного вектора сигналов) и корреляционной матрицы, заданной весовыми коэффициентами корреляции. Процесс изменения состояний системы является итеративным и определяется рекуррентными связями. Конечное состояние системы характеризуется достижением максимально возможной (устойчивой) корреляции реакций нейронов (их состояний) в соответствии с их весовыми коэффициентами. Энергетическая функция (функционал Ляпунова) в данной интерпретации приобретает демонстрационный смысл и отображает стремление системы к стабильному состоянию или уменьшению внутренней информационной энтропии.

Однако, если мы решили рассмотреть работу этой сети исключительно с энергетической точки зрения, то давайте обратим внимание на то, что внутренняя энергия сети, как мы установили, является энергией нейронов (юнитов) в соответствии с формулами (13, 14), тогда как энергия взаимосвязи нейронов остается неопределенной. Представим, что система "не знает", что такое бит информации, а рассматривает его как значение потенциальной энергии, которую нужно "куда-то деть": сохранить, распределить, рассеять или преобразовать в энергию выхода. Тогда динамика изменения состояний нейронов определяется

лишь элементами внешней энергии, поступающей в систему. В нашем случае - в виде энергии, соответствующей символам x_i биполярного (+1 - максимальная энергия, или -1 - минимальная энергия) алфавита системы. Действительно, если мы не подаем сигналы на вход сети, то на выходе каждого нейрона у нас будут либо нулевые, либо случайные значения энергии. Таким образом, значения внутренней энергии сети, выраженные в виде матрицы весовых коэффициентов w_{ij} , определяют пороговые значения состояния покоя системы. Динамика изменения состояний системы определяется превышением внешней энергии над этими пороговыми значениями покоя в пределах порогового значения активации θ_i . Этот процесс нужно рассматривать не как уменьшение внутренней энергии системы до минимального (стабильного) уровня, а как ее перераспределение между локальными точками энергетического ландшафта (юнитами). Важной особенностью модели является и то, что энергетический ландшафт, определяемый матрицей весовых коэффициентов, формируется на основе образцов внешней энергии, которые необходимо запомнить, а не в виде условной энергии ошибки, как, например, в многослойном перцептроне. Тогда, энергетический ландшафт модели Хопфилда можно было бы рассматривать как *внутреннюю модель внешнего мира, воспринимаемого системой*.

Таким образом, процесс итерационного перераспределения внешней энергии по локальным уровням энергетического ландшафта представлен в виде функционала Ляпунова, выраженного функцией энергии Хопфилда (19). Результат этого перераспределения показывает уровень превышения внешней энергии над внутренней энергией состояния покоя в точках локальных экстремумов ландшафта, который, после его сравнения с пороговым значением θ_i функционала, отображается в выходных сигналах системы. Тогда, θ_i - это порог принятия решения для перераспределения внешней энергии между элементами сети. Согласно формуле Хопфилда (19), с каждым шагом перераспределения энергии, общая условная энергия сети E уменьшается, что в соответствии с законами

термодинамики отражает процесс увеличения термодинамической энтропии данной системы, т.е. внешняя энергия уменьшается в результате ее "рассеивания" между нейронами и заполнения "ям" потенциальной энергии юнитов (нейронов), пока не будет достигнуто состояние, в котором дальнейшие изменения невозможны и система становится стабильной. Тогда, мы также могли бы сделать вывод, что формула энергии Хопфилда – это формула динамики изменения энтропии, которая по своей сути связывает информационную и термодинамическую энтропию. Однако, недостатком этого описания является то, что нейроны должны "рассеять" потенциальную энергию, скопившуюся в его "яме" до следующего цикла обновления состояния. Таким образом, система для поддержания динамики работы будет постоянно требовать дополнительной внешней энергии.

Одной из концептуальных проблем классической модели Хопфилда и данного альтернативного энергетического описания ее работы является отсутствие механизма самоостановки. Сама система будет работать бесконечно без изменения состояний (в режиме обновления) пока не придет внешний сигнал остановки.

Таким образом, сеть Хопфилда как энергетическую и информационную систему можно классифицировать как систему с *детерминированной статической структурой, детерминированным статическим обобщенным энергетическим ландшафтом в состоянии покоя, который формируется инстантно (не итеративно), и детерминированным динамическим энергетическим ландшафтом в процессе минимизации информационной энтропии.*

4. Энергетическая модель формального нейрона.

Модель нейрона МакКаллока-Питтса (МКП) является упрощенной математической моделью биологического нейрона, которая, тем не менее, получила широкое распространение и легла в основу построения современных

ИНС. Эта информационная модель реализуется алгоритмически, однако моделирование ее как энергетической системы позволит понять какие ограничения имеет данная модель по сравнению с реальной физической системой, которой безусловно является и биологический нейрон.

Рассмотрим модель МКП как однослойный перцептрон Розенблатта в ее классическом варианте с бинарными сигналами $\{0, 1\}$, значениями весовых коэффициентов в диапазоне $\{-1, +1\}$ и активационной функцией Хевисайда. Энергетическая модель МКП может быть рассмотрена на основе построения ее энергетического ландшафта. Мы уже говорили о том, что энергетический ландшафт нейрона МКП в состоянии покоя представлен множеством весовых коэффициентов, которые определяют энергетический вклад каждого входа в преодоление "потенциальной энергетической ямы", глубина которой ограничена сверху порогом активации Q (12-16). Также, мы уже говорили о том, что энергия модели МКП представляет собой только энергию юнита энергетического ландшафта более сложной системы (например ИНС). Данный юнит (нейрон) не формирует энергию взаимосвязи с другими юнитами, в отличие от реальных физических систем (9). На Рис.2 представлена общая структура энергетической модели МКП, где $X = (x_1, x_2, \dots, x_n)$ – множество входных сигналов, представляющих входную энергию системы; Y – реакция модели МКП, представляющая выходную энергию системы; $W = (w_1, w_2, \dots, w_k)$ – весовые коэффициенты, представляющие энергетические вклады входов в преодоление порога активации; Σ – сумматор взвешенных входов, представляющий текущую глубину "потенциальной ямы" системы (юнита); f – активационная функция Хевисайда, представляющая верхний порог "потенциальной ямы".

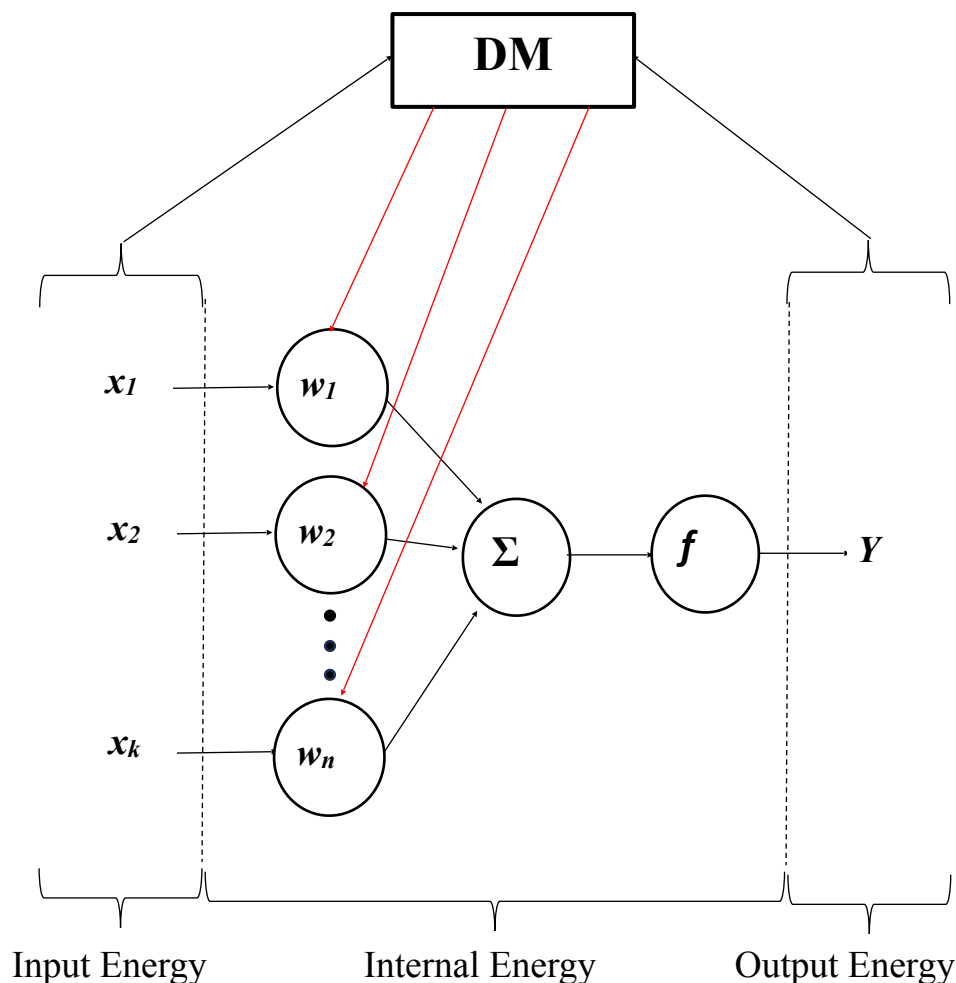


Рис.2. Общая структура энергетической модели МКП

Сумма взвешенных входов представляет внутреннюю энергию системы. Эта энергия делится на возбуждающую (энергию активации), где весовые коэффициенты имеют знак "плюс", и ингибирующую (тормозящую) энергию, где весовые коэффициенты имеют знак "минус".

Таким образом, если "положительная" суммарная энергия входов превышает "отрицательную" или тормозящую энергию, то порог активации $Q = 0$ превышает и система, используя активационную функцию, принимает решение о формировании выходного сигнала $Y = 1$ вне зависимости от значения V – потенциала активации

$$V = \sum_i w_i^+ x_i + \sum_i w_i^- x_i = V^+ + V^- \quad (23),$$

где w_i^+ - положительные весовые коэффициенты, определяющие положительный потенциал активации V^+ , w_i^- - отрицательные весовые коэффициенты, определяющие тормозящий потенциал V^- . Тогда, именно тормозящий потенциал V_{rest} определяет "глубину" потенциальной ямы и минимальную энергию юнита в состоянии покоя V_{rest}

$$V_{\text{rest}} = V_{\text{rest}}^- = \sum_i w_i^- \quad (24).$$

Эта "глубина" потенциальной ямы представлена на энергетическом ландшафте модели юнита (формального нейрона) в состоянии покоя (Рис. 3). Процесс нормализации энергии на выходе системы (представление ее в виде единичного значения бита на основе принципа Ландауэра) может быть связан как с "рассеиванием" энергии превышающей нормализованное значение "1", так и с использованием дополнительной энергии в случае $V < 1$.

Если же $|V^-| > |V^+|$, то вся энергия V^+ будет рассеяна с выделением энергии в соответствии с принципом Ландауэра (7).

В модели на Рис.2 также представлен Демон Максвелла (DM) [18]. Это абстрактная термодинамическая концепция, предложенная шотландским физиком Джеймсом Клерком Максвеллом, которая широко используется в физике информации и квантовых вычислениях. В данном случае мы рассмотрим ее как аналогию с внешней энергетической системой, по отношению к рассматриваемой системе, которая реализует алгоритм обучения нейрона. DM, как энергетическая

система, регулирует поток энергии входов для его преобразования во внутреннюю энергию системы, целенаправленно меняя значения весовых коэффициентов.

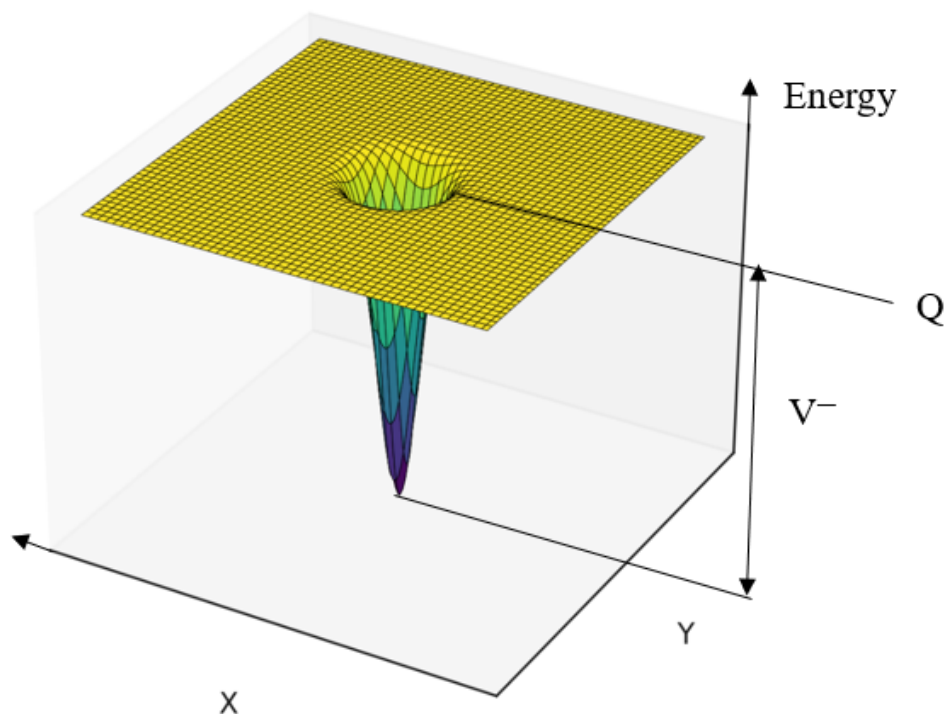


Рис.3. Энергетический ландшафт модели МКП в состоянии покоя

Практически, DM минимизирует энергию ошибки на выходе системы. Эта энергия ошибки представляет собой "псевдоэнергию" по отношению к внутренней и внешней энергии системы.

Для рассмотрения модели МКП как термодинамической системы, нам необходимо также определить микро и макросостояния системы. Под *микросостояниями* системы мы будем понимать *множество взвешенных входов системы*, которые будут соответствовать определенным *выходам системы* – *макросостояниям*. В данной модели у нас будет всего два макросостояния системы $Y = \{0, 1\}$, причем, только макросостояние $Y = 1$ соответствует активации нейрона.

Рассмотрим вероятность активации нейрона P_{act} как условную вероятность

$$P_{act} = \sum_{V^-} P(V^-) P(V^+ \geq (Q + V^-) | V^-) \quad (25),$$

где $P(V^-)$ – вероятность того, что тормозящие входы дадут суммарное значение V^- ;

$P(V^+ \geq (Q + V^-) | V^-)$ – условная вероятность того, что возбуждающие входы суммарно дадут значение, превышающее эффективный порог $(Q + V^-)$.

Это отражает влияние тормозящих вкладов: чем "глубже" (т.е. больше по модулю V^-), тем выше эффективный порог для возбуждающих входов, а значит, тем меньше вероятность, что условие $V^+ \geq (Q + V^-)$ выполнится.

Если мы определяем число микросостояний, которые приводят к активации, как Ω_{act} , то термодинамическая энтропия S_{therm} будет определяться формулой

$$S_{therm} = k_B \ln \Omega_{act} \quad (26).$$

Количество микросостояний Ω_{act} будет зависеть именно от того, сколько комбинаций взвешенных возбуждающих и тормозящих входов удовлетворяют условию $V^+ \geq (Q + V^-)$. Т.е. более "глубокая яма" (сильное торможение) уменьшает Ω_{act} так как требуется больше возбуждения для преодоления порога, что ведет к снижению термодинамической энтропии для данного макросостояния активации.

Аналогичные рассуждения справедливы и для второго макросостояния $Y = 0$, где число микросостояний, которые не приводят к активации $\Omega_{non-act}$, будут определяться при выполнении условия $V^+ < (Q + V^-)$.

При равновероятном распределении всех 2^N микросостояний, где N – число входов, можно записать вероятность активации как

$$P_{act} = \frac{\Omega_{act}}{2^N} \quad (27).$$

Тогда

$$\Omega_{act} = 2^N P_{act} \quad (28).$$

В этом случае, термодинамическая энтропия для макросостояния активации S_{act} , будет равна

$$S_{act} = k_B \ln(2^N P_{act}) = k_B (N \ln 2 + \ln P_{act}) \quad (29).$$

Токаим образом, при увеличении P_{act} увеличивается и термодинамическая энтропия при фиксированном значении N и равновероятном распределении микросостояний.

Аналогичные рассуждения справедливы и для определения термодинамической энтропии для второго макросостояния $S_{non-act}$.

Однако, в процессе обучения нейрона мы ограничиваем количество допустимых комбинаций входов и, следовательно, делаем их неравновероятными для заданного (определенного) макросостояния, т.е. мы, например для активации нейрона, применяем условную вероятность (25). В этом случае, мы используем только значимое подмножество T из всего множества комбинаций 2^N .

$$|T| \ll 2^N \quad (30).$$

Если мы считаем, что внутри T распределение комбинаций равномерное, то

$$P_{act} = \frac{\Omega_{act}}{|T|} \quad (31),$$

что, при сохранении той же вероятности P_{act} в условиях неравномерного распределения комбинаций входов, которое ведет к выделению значимого подмножества T , требует меньшее число входных комбинаций (эффективно активных), и, следовательно, меньшее число возможных микросостояний Ω_{act} . То же справедливо и для подмножества комбинаций T' , которые не ведут к активации нейрона.

Таким образом, мы видим, что при обучении нейрона его термодинамическая энтропия, как энергетической системы, уменьшается для выделенных значимых подмножеств входных комбинаций T и T' , в то время как для общего равновероятного распределения комбинаций входов, термодинамическая энтропия для двух макросостояний будет увеличиваться в результате увеличения количества микросостояний до значения 2^N .

Рассмотрим информационную энтропию (Шенноновскую) H модели МКП

$$H = - [P_{act} \log_2 P_{act} + (1 - P_{act}) \log_2 (1 - P_{act})] \quad (32).$$

Мы видим, что H достигает максимума при $P_{act} = 0.5$, т.е. при равновероятностном распределении комбинаций взвешенных возбуждающих и тормозящих входов, что соответствует и максимуму термодинамической энтропии.

При условной вероятности P_{act} (25), приближающейся к 0 или к 1, мы видим уменьшение информационной энтропии, что также соответствует локальному уменьшению термодинамической энтропии для соответствующего макросостояния.

Таким образом, если мы определяем информационную энтропию через распределение вероятностей по микросостояниям, то оказывается, что она точно соответствует термодинамической энтропии (с учётом константы k_B), а процесс обучения, который ведёт к селекции значимого подмножества входных комбинаций, изменяет распределение микросостояний, что и приводит к локальному уменьшению термодинамической энтропии.

Теперь давайте рассмотрим энергию выхода системы как свободную энергию Гиббса G [19]

$$G = H - TS \quad (33),$$

где H – энтальпия системы (теплосодержание, энергия, включая теплоту);

T – абсолютная температура; S – термодинамическая энтропия системы.

Свободная энергия Гиббса — это термодинамическая функция состояния, которая характеризует способность системы совершать полезную работу при изотермических и изобарных условиях (то есть при постоянной температуре и давлении). Этот термин широко используется в химической термодинамике для анализа энергетических аспектов химических реакций и фазовых переходов, но его интерпретация как энергии на выходе системы не является широко используемой. Как правило, это показатель того, сколько энергии может быть превращено в полезную работу: механическую, электрическую или другую. Важно то, что свободная энергия Гиббса может быть связана с внутренними структурами системы (например, кристаллическая решётка, молекулярная организация). В этом случае значение G отражает, как структура системы влияет на её способность выполнять работу. Тем не менее, в машинном обучении и байесовской статистике применяется метод вариационной свободной энергии, который использует термодинамические принципы для оптимизации [20]. Здесь вводится функционал,

аналогичный энергии Гиббса, чтобы минимизировать разницу между моделью и данными.

Однако, если мы представим, что свободная энергия на выходе информационной системы "выполняет работу" по созданию внутренней модели внешнего мира для другой системы, которая ее воспринимает, то аналогия информации, которая формируется на выходе системы, и свободной энергии Гиббса приобретает совсем другой смысл и становится обоснованной. Кроме того, информация позволяет системам принимать решения, прогнозировать или изменять своё состояние. Например, в системах управления или робототехнике выход ИНС используется для управления реальными физическими объектами и это является аналогом физической работы.

Тогда,

$$G = E_{str} = E_{total} - E_{unstr} \quad (34),$$

где E_{str} – структурированная часть энергии на выходе системы. Эта часть энергии содержит измеряемые параметры, которые отражают упорядоченность, связанную с внутренней структурой системы. Она связана с уменьшением информационной энтропии благодаря структурированию данных (например, классификация, выделение признаков, предсказания и т. д.);

E_{total} – общая энергия, выделяемая системой (энтальпия);

E_{unstr} – неструктурированная часть энергии. Эта часть энергии остаётся в форме, не несущей устойчивой структуры энергии на выходе системы, которая может быть представлена как информации о внутренней структуре системы. Эта энергия аналогична тепловым потерям или шуму, увеличивающему термодинамическую и информационную энтропию.

Мы видим, что для модели МКП только для макросостояния "1" происходит формирование (выделение) свободной энергии, а для макросостояния "0", которое

определяет отсутствие реакции нейрона, вся энергия представляет собой термодинамические потери (TS) в соответствии с принципом Ландауэра. Если не принимать во внимание энергетические затраты при нормализации единичного выхода нейрона: дополнительную энергию при $0 \leq V < 1$, или дополнительное рассеивание тепла при $V > 1$, то можно считать, что тормозящий потенциал V вносит весомый вклад в формирование $TS = E_{unstr}$

$$TS = \Omega V^{-E} \quad (35),$$

где E – энергия Ландауэра (7), определяющая тепловую энергию, выделяющуюся при стирании одного бита информации, Ω - общее число микросостояний, V - тормозящий потенциал, определяющий число битов, которые стираются для одного микросостояния. Тогда, общая неструктурированная часть выделяемой энергии, как "хаотическая" часть энтальпии, равна общей термодинамической энтропии нейрона МКП

$$S = \Omega V^{-k_B \ln 2} = 2^N V^{-k_B \ln 2} \quad (36).$$

Мы видим, что общая термодинамическая энтропия модели МКП растет по мере увеличения количества микросостояний и "глубины" потенциальной ямы нейрона (тормозящего потенциала).

Следовательно, с увеличением общего количества микросостояний системы растет энтальпия системы и растет ее неструктурированная составляющая E_{unstr} , в то время как структурированная часть энергии выхода E_{str} достигает своего максимального значения для ограниченного числа микросостояний после обучения, которое соответствует минимуму термодинамической и информационной энтропии для соответствующих макросостояний, и в дальнейшем остается неизменной.

Этот вывод согласуется с выводами теории диссипативных систем Пригожина, где говорится о том, что система, находясь далеко от равновесия, может "обучаться", то есть переходить в состояние, которое является аттрактором динамики. Это означает, что из огромного числа возможных микросостояний система "выбирает" те, которые оптимально поддерживают упорядоченность (низкую внутреннюю энтропию) за счёт диссипации энергии. Таким образом, хотя общая энтропия окружающей среды может продолжать расти, внутренняя структура системы стабилизируется.

Уточним, что в классической равновесной термодинамике свободная энергия Гиббса минимизируется. Здесь же мы имеем дело с открытой системой, которая, с учетом "Демона Максвелла", только условно может рассматриваться как самоорганизующаяся система. В этом случае понятие "максимума" структурированной энергии E_{str} следует понимать в терминах оптимального функционального состояния системы, а не как классическое равновесное состояние. Это не противоречит и классическому пониманию свободной энергии Гиббса, т.к. мы видим, что с ростом общей энтальпии системы растет и общая термодинамическая энтропия системы (E_{unstr}), в то время как свободная энергия (E_{str}), а вернее ее доля (составная часть энтальпии) – уменьшается, как уменьшается и ее локальная термодинамическая энтропия.

В целом, эти выводы согласуются с *динамическим уравнением эволюции системы*, которое является аналогом уравнения Ланжевена в статистической механике, описывающего движение частицы в потенциальном поле с тепловым шумом

$$\frac{dx}{dt} = - \frac{\partial V(x)}{\partial x} + \eta(t) \quad (37),$$

где x – внутренняя энергия системы; $V(x)$ — эффективный потенциальный ландшафт энергии, $\eta(t)$ - стохастический шум, моделирующий тепловые флуктуации, связанные с термодинамическими потерями [21, 22].

В данном уравнении функция $V(x)$ играет роль эффективного энергетического потенциала, определяющего, как система стремится минимизировать свою внутреннюю энергию. Тогда, $-\partial V(x)/\partial(x)$ будет определять естественную тенденцию системы снижать свою энергию, т.е. переходить к более устойчивым состояниям.

Если в левую часть уравнения добавить коэффициент диссипации γ , то получим уравнение для диссипативной системы, которое напоминает уравнение релаксации энергии в открытых системах.

Общие выводы:

1. В модели МКП, как и в моделях ИНС, информация представлена в виде абстрактных элементов внутреннего алфавита системы, например, битов. Эти элементы алфавита не имеют внутренней структуры, которую мы могли бы ассоциировать со структурой энергией, как, например, структура энергии элементов в квантовой физике или химии. Однако информационные биты связаны с энергией в соответствии с принципом Ландауэра, что позволяет рассматривать энергетическую модель как нейрона, так и в целом ИНС. В этом случае, внутренняя энергия модели МКП, как юнита энергетической системы, может быть рассмотрена только на основе внешней "псевдоэнергии", параметры которой (весовые коэффициенты) формируются "Демоном Максвелла" (внешним алгоритмом обучения) на основе статистических наблюдений. Статистические параметры внешней "псевдоэнергии" для определения внутренней энергии юнита (модели МКП) могут рассматриваться только как "поверхностные" свойства энергии системы, которые не в полной мере отображают структуру внутренней энергии юнита и не могут использоваться для формирования энергии взаимосвязи юнитов, как это происходит в физических системах.

2. В процессе обучения модели МКП осуществляется формирование устойчивой структурированной части излучаемой (генерируемой) системой энергии, которая может рассматриваться как свободная энергия Гиббса. Эта структурированная энергия имеет минимальное значение термодинамической и информационной энтропии для ограниченного числа эффективных микросостояний, которые соответствуют определенным (заданным в процессе обучения) макросостояниям. При этом, динамика уменьшения обеих энтропий одинакова, что отображает глубокую энергетическую природу их взаимосвязи. В процессе обучения модель МКП, а вернее "Демон Максвелла" как ее часть, минимизирует внешнюю "псевдоэнергию" в виде ошибки реакции модели. Эта минимизация отображается в изменении статистических параметров внутренней энергии юнита, которые представлены в виде весовых коэффициентов. Это ведет к сокращению числа эффективных микросостояний и, в конечном итоге, к минимизации внутренней энергии системы.

3. Основываясь на приведенной интерпретации свободной энергии, можно сказать, что параметры структурированной энергии на выходе системы (для модели МКП – это может быть последовательность двоичных сигналов) отображают внутреннюю энергию системы, а значит и ее внутреннюю структуру, которая может интерпретироваться как *внутренняя модель восприятия мира*.

5. Энергетические модели многослойных ИНС

Для понимания энергетической природы информации рассмотрим многослойные ИНС, которые можно считать идеальными моделями не только для изучения принципов обработки информации, но и для изучения их взаимосвязи с энергией физических систем. С этой целью воспользуемся концепцией энергетического ландшафта и рассмотренной энергетической моделью МКП.

На Рис. 4 представлена модель энергетического ландшафта многослойного перцептрона (MLP).

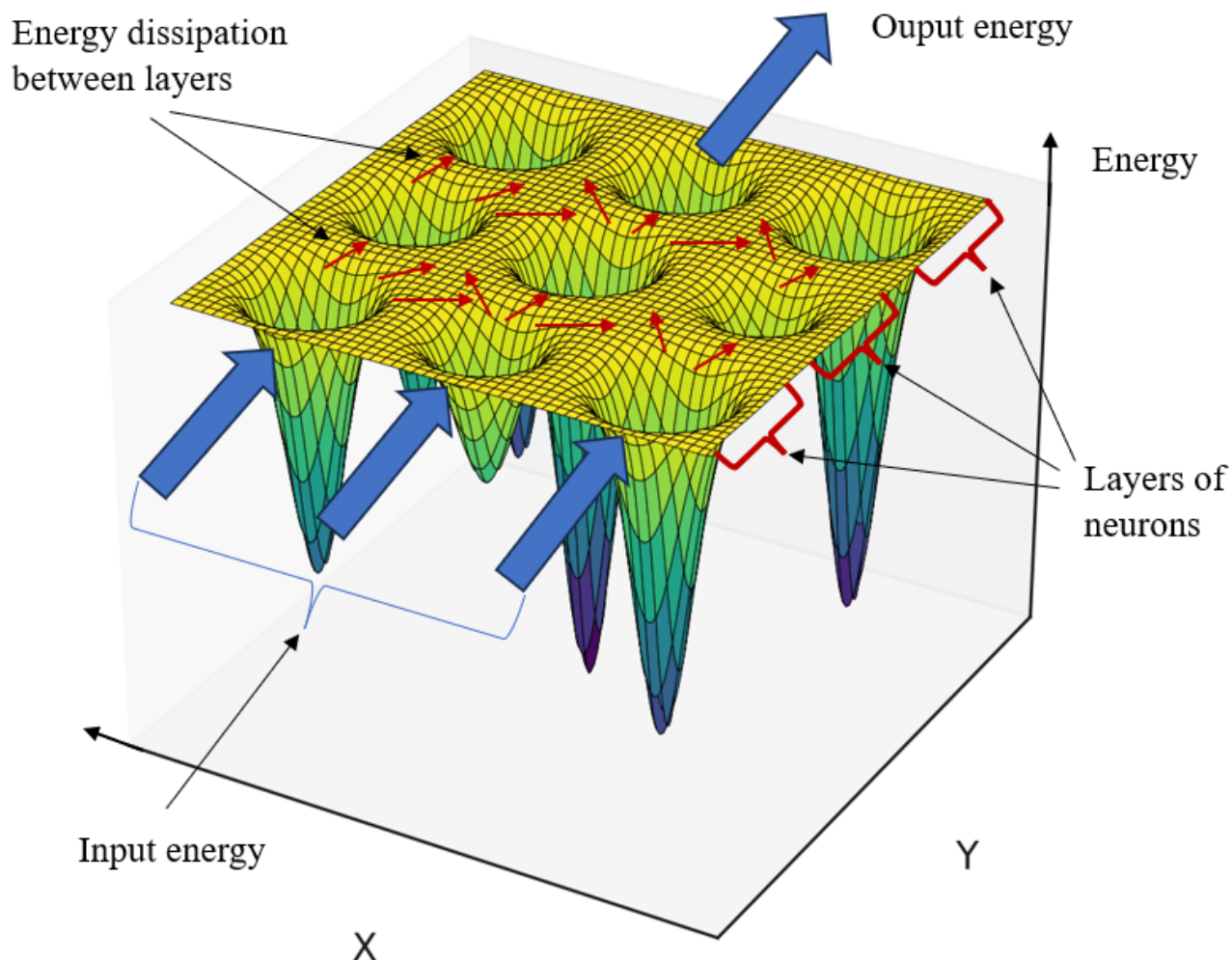


Рис.4. Модель энергетического ландшафта MLP

Этот энергетический ландшафт состоит из потенциальных "ям" юнитов (нейронов МКП) разной глубины в состоянии покоя, которые образуют скрытые слои. Модель иллюстрирует общую концепцию конвергенции энергии, которая отображает процесс конвергенции информации в MLP при решении задачи классификации. Основной особенностью этого ландшафта является то, что на нем отсутствуют энергетические взаимосвязи – каналы распределения энергии между

юнитами. Выходная энергия юнитов просто рассеивается между входами юнитов следующего слоя. Это является следствием того, что весовые коэффициенты входов не формируют энергию взаимосвязей, а лишь определяют глубину потенциальной "ямы" юнита. Процесс рассеивания выходной энергии юнитов представлен в виде полносвязной архитектуры слоев MLP. В реальной энергетической системе такое рассеивание потребовало бы усиления информационного сигнала на каждом входе нейрона. Отсутствие энергии взаимосвязи между структурными элементами в данной энергетической модели является ограничением по сравнению с реальными энергетическими системами (Рис.1.). Согласованное формирование энергии юнитов (нейронов) в этой модели ИНС (т.е. формирование энергетического ландшафта системы в состоянии покоя) также связано с процессом минимизации "превдоэнергии", которым управляет "Демон Максвелла" этой системы (например, алгоритм обратного распространения ошибки). Этот процесс формирования согласованных энергий юнитов для достижения множества требуемых макросостояний системы является компромиссом между общностью формируемых признаков классификации и точностью решений этой задачи для заданной группы классов. Для многих практических задач, достижение этого компромисса в процессе обучения не позволяет добиться идеального решения.

Основываясь на результатах исследований, представленных в разделе 4, можно сделать вывод о том, что информационная и термодинамическая энтропии системы, энергетический ландшафт которой состоит из множества отдельных энергетических юнитов (моделей МКП), будут демонстрировать те же свойства, что и отдельный юнит.

Доказательством этих выводов являются результаты исследований Ravid Shwartz-Ziv and Naftali Tishby (2017) и др. [23, 24], в которых авторы предложили теорию Information Bottleneck (IB). Эта теория рассматривает глубокие нейронные сети, такие как MLP, и демонстрирует, что в процессе их обучения наблюдается

тенденция к уменьшению информационной энтропии в реакциях нейронов скрытых слоев. Кроме того, уменьшение информационной энтропии сопровождается снижением вариативности реакций нейронов в скрытых слоях, что свидетельствует об уменьшении числа микросостояний для рассматриваемых реакций (макросостояний) системы. Это также свидетельствует о локальном снижении термодинамической энтропии для этих макросостояний.

Так, в работе Shwartz-Ziv and Tishby (2017) [23] показано, что в задачах классификации на наборах данных, таких как MNIST, информационная энтропия активации падает в 2-4 раза от входного слоя к выходному, что соответствует снижению дисперсии активации от 50% до 90% между начальным и глубокими слоями.

В теории IB авторы вводят понятие взаимной информации: $I(X; T)$ – информация, которую слой T содержит о входных данных X ; $I(Y; T)$ – информация, которую слой T содержит о целевых данных Y . Взаимная информация $I(X; Y)$ определяет насколько знание одной переменной уменьшает неопределенность другой. В общем случае, взаимная информация равна сумме энтропий отдельных переменных за вычетом их совместной энтропии

$$I(X; Y) = H(X) - H(X|Y) \quad (38);$$

$$I(X; Y) = H(Y) - H(Y|X) \quad (39);$$

$$I(X; Y) = I(Y; X) \quad (40);$$

$$I(X; Y) = H(X) + H(Y) - H(X, Y) \quad (41).$$

Если X и Y независимы, то знание Y не уменьшает неопределённость X , тогда

$$H(X|Y) = H(X) \text{ and } I(X; Y) = 0 \quad (42).$$

Если X и Y полностью зависимы, то знание одной переменной полностью определяет другую, и $I(X; Y)$ достигает максимального значения, равного энтропии той переменной, которая имеет меньшую энтропию.

Эти исследования показывают, что $I(X; T)$ значительно уменьшается к последним слоям, что указывает на значительное снижение неопределенности входных данных, а информационная энтропия активации $H(T)$ в каждом слое уменьшается в процессе обучения. Например, для сети с 10 слоями на последнем слое $H(T)$ снижается примерно на 50-70% по сравнению с первым слоем. Однако, как и следовало ожидать, на выходе системы сохраняется информационная энтропия, что подтверждает существование компромисса между общностью формируемых признаков классификации, которая обусловлена обобщенным энергетическим ландшафтом системы, и стабильностью (точностью) реакций системы на определенную входную информацию.

Также было показано, что для начальных слоев на старте обучения взаимная информация между представлением слоя и входом $I(T; X)$, может достигать порядка 10–15 бит. Это означает, что на этих слоях сохраняется большое количество исходной (сырой) информации. Для средних слоев, по мере прохождения информации, значения $I(T; X)$ начинают снижаться. Например, уже во втором по глубине слое $I(T; X)$ может упасть до 8-10 бит. В последних скрытых слоях наблюдается наиболее выраженное сжатие информации. Здесь $I(T; X)$ может снизиться до 2–3 бит, при этом $I(T; Y)$ остаётся близким к энтропии меток, то есть все релевантные информационные признаки, необходимые для решения, сохраняются.

Эти практические результаты полностью подтверждают наши теоретические выводы об уменьшении информационной энтропии и локальном уменьшении структурированной части термодинамической энтропии на выходе рассматриваемой системы (энергетической модели ИНС) для определенного множества ее макросостояний.

Таким образом, так как рассматриваемая система не имеет энергии взаимосвязи между юнитами, то ее внутренняя энергия представлена исключительно энергией юнитов. Каждый юнит воспринимает энергию соседнего юнита как внешнюю энергию, преобразуя ее в свою внутреннюю энергию с помощью весовых коэффициентов, которые играют роль своеобразных детекторов входной энергии. Тогда, активационную функцию нейрона мы можем интерпретировать как функцию преобразование внутренней энергии в выходную энергию юнита. Структурированная внешняя энергия всей системы, отображающая информацию на выходе ИНС, формируется активационными функциями нейронов внешнего (конечного) слоя. Эта система реализует процесс параллельного восприятия энергии (информации) и ее конвергенции (сжатия) в процессе обучения.

Назовем такую энергетическую модель ИНС – моделью первого типа. К этому типу ИНС также относится сверточная сеть (CNN), основное отличие которой заключается в сегментации входного вектора сигналов (входной энергии) и локальной обработке отдельных сегментов.

Модель ИНС первого типа можно классифицировать как энергетическую систему с детерминированной статической структурой, детерминированным статическим обобщенным энергетическим ландшафтом в состоянии покоя, который формируется итеративно "Демоном Максвелла", и детерминированной динамикой изменения энергии юнитов на энергетическом ландшафте в процессе минимизации информационной энтропии и локальной минимизации термодинамической энтропии.

Мы также можем сделать вывод, что обобщенный статический ландшафт системы в состоянии покоя, представленный исключительно энергией юнитов, и детерминированная статическая структура системы являются концептуальным ограничением ее точности при увеличении заданного количества макросостояний системы.

В текстовых генеративных моделях ИНС, таких как GPT, в отличие от моделей ИНС первого типа, происходит дивергенция входной энергии и информации, которую мы ассоциируем с "предсказаниями". Эти системы имеют совершенно другой энергетический ландшафт, который представлен не энергией юнитов, как, например, в моделях MLP, а энергией взаимосвязи между юнитами, в то время, как сами юниты представлены только статическими элементами внутреннего алфавита системы. Для генеративных моделей ИНС этими элементами являются, например, слова естественного языка (или их составные части), представленные в виде токенов – их внутренних числовых представлений (эмбеддингов).

В общем случае статические эмбеддинги, как например в моделях Word2Vec или GloVe, представляют слова в виде векторов фиксированной длины в высокоразмерном пространстве, где их близость (например, по косинусной схожести) может служить косвенным индикатором сходства или ассоциации между словами. Эту ассоциативную связь мы также можем условно интерпретировать как "вероятность взаимосвязи" между словами, что не является прямым пониманием вероятности в математическом смысле. Эта взаимосвязь отражает только усреднённые статистические соотношения из обучающего корпуса слов. Такой вектор можно рассматривать как репрезентацию "центрального смысла" слова, но он не учитывает многозначность или изменение значений в разных контекстах.

Тогда как контекстуальные эмбеддинги в моделях на основе трансформеров, например, BERT или GPT, могут преобразовывать базовое представление слова - lookup вектор с учётом окружающих слов - контекста. Этот процесс отображает динамическое выделение тех аспектов значения слова, которые актуальны в данном контексте. Одно и то же слово может иметь разные векторы в зависимости от синтаксиса, семантики и даже стиля текста, что позволяет модели учитывать его полисемию. Это позволяет динамически выбирать, а вернее вычислять на основе

матрицы весов механизма внимания (query-key-value), наиболее релевантные (наиболее вероятные, при использовании активационной функции типа *Softmax*) для данного контекста взаимосвязи слов. Эта динамика изменения взаимосвязей в процессе инференса реализуется механизмами Multi-Head Attention в различных слоях модели.

Все взаимосвязи слов (эмбединги) формируются на основе матриц весовых коэффициентов, которые, в свою очередь, формируются в процессе обучения (в энергетической интерпретации – "Демоном Максвелла"). Это означает, что все потенциальные пути взаимосвязей между словами "заложены" в матрицах весовых коэффициентов модели, поскольку именно они формируют структуру возможных ассоциаций и преобразований. Однако конкретные взаимосвязи, актуальные для данного контекста, выбираются (вычисляются) динамически при обработке входного текста. Таким образом, матрицы весовых коэффициентов модели формируют глубину "потенциальных каналов" взаимосвязей между юнитами энергетического ландшафта системы в состоянии покоя, а активационные функции нейронов модели определяют "выходную энергию" из этих "каналов", которая и активирует "статические энергии" юнитов. Эти "энергии" юнитов и формируют структурированную часть выходной энергии системы. Тогда, нейроны генеративных моделей практически реализуют взаимосвязи между элементами внутреннего алфавита системы (юнитами).

Упрощенная модель энергетического ландшафта генеративной ИНС представлена на Рис.5.

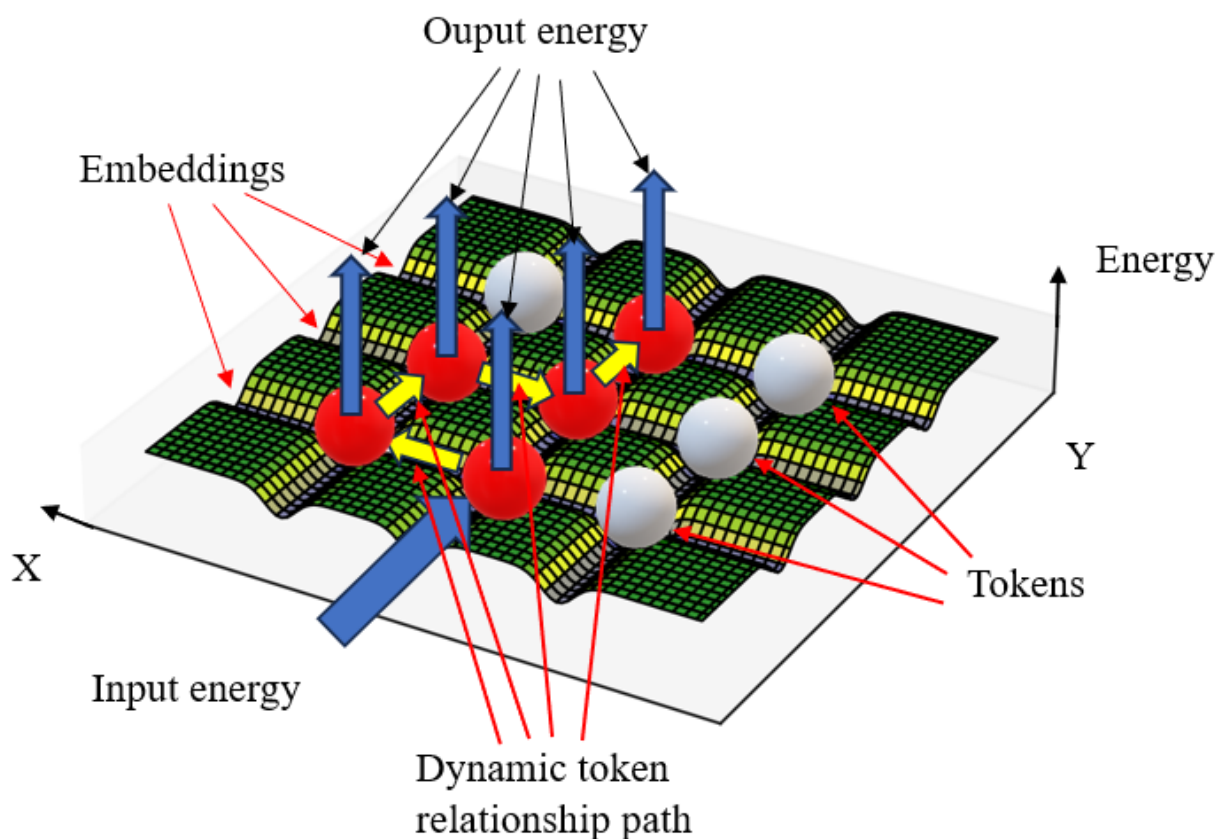


Рис.5. Модель энергетического ландшафта генеративной ИНС

Таким образом, эта система реализует процесс последовательного восприятия потока энергии (информации) и ее дивергенции (расхождения) по различным "каналам" энергетического ландшафта системы с целью активации условной энергии юнитов (токенов).

Назовем такую энергетическую модель ИНС – моделью второго типа. К этому типу могут относиться рекуррентные ИНС, а также генеративные модели на основе трансформеров.

Эта модель имеет сложный энергетический ландшафт, сформированный множеством матриц весовых коэффициентов различных видов (в моделях трансформеров используется около 8 типов матриц для реализации механизмов внимания, эмбединга, линейных трансформаций в полносвязных слоях, нормализации и формирования выхода), а также множеством скрытых слоев и

нейронов в каждом слое (например, в модели GPT-3 размер скрытого слоя может быть до 12288 нейронов для самой крупной модели с 96 слоями), что определяет огромное количество параметров модели (например, в модели GPT-3 содержится до 175 миллиардов параметров). Эти параметры и их комбинации формируют множество возможных микросостояний системы. Тогда, динамика выбора наиболее релевантной взаимосвязи токенов (юнитов) в процессе инференса (работы обученной генеративной ИНС) определяет сокращение пространства микросостояний до минимально возможного (в зависимости от механизма альтернативного выбора токенов). В идеале - до одного фиксированного множества микросостояний, которое и будет определять макросостояние системы в виде выходной информации.

Таким образом, генеративная модель ИНС в процессе обучения увеличивает пространство возможных микросостояний системы (формирует его), а в процессе инференса – сокращает его до минимума, что и формирует макросостояния системы. В то же время, модель первого типа сокращает пространство микросостояний в процессе обучения для заданного макросостояния.

Очевидно, что информационная энтропия на выходе ИНС второго типа также уменьшается. Это связано с тем, что генеративные модели в процессе инференса выбирают наиболее релевантные взаимосвязи, т.е. наиболее вероятную контекстную структуру токенов. Этот процесс сокращает число возможных комбинаций токенов, что и уменьшает информационную энтропию сгенерированного текста, в идеале - до одной структуры для определенной модели (хотя на практике, с целью максимизации вероятности последовательности токенов может использоваться эвристический алгоритм Beam Search для поиска нескольких наиболее высоковероятных вариантов их взаимосвязей, что приводит к более предсказуемым, но менее разнообразным результатам). Это поведение можно рассматривать как аналог эффекта "бутылочного горлышка" в ИНС первого типа. Для увеличения разнообразия текстов (ответов) в генеративной модели

(например, GPT-4), наоборот, приходится искусственно увеличивать информационную энтропию на выходе системы, применяя такие стохастические методы, как температурное сглаживание (масштабирование), которое добавляет случайность в выбор токенов, или методы Top-k/Top-p сэмплинга, которые ограничивают набор кандидатов для выбора следующего токена, что вводит дополнительный элемент случайности и позволяет генерировать более разнообразные тексты. Из-за этого пространство микросостояний не схлопывается до единственного варианта, а остаётся вероятностным, что и даёт вариативность в ответах модели.

Таким образом, статистическая природа обучения генеративных ИНС приводит к тому, что на выходе мы получаем наиболее вероятную, т.е. контекстно усредненную структуру токенов. Это является естественной причиной уменьшения информационной энтропии на выходе системы. Однако, как уже обсуждалось, наиболее информативной и потенциально ценной для решения проблемы является именно взаимосвязь с низкой вероятностью, которая отображает новую или "креативную" информацию.

Мы также видим, что уменьшение количества возможных микросостояний системы в процессе инференса, которое ведет к формированию одного макросостояния системы на выходе, свидетельствует о локальном уменьшении структурированной части термодинамической энтропии системы (34).

Таким образом, модель ИНС второго типа можно классифицировать как энергетическую систему с *детерминированной статической структурой, детерминированным статическим обобщенным энергетическим ландшафтом в состоянии покоя, который формируется итеративно "Демоном Максвелла", и детерминированной динамикой изменения энергии взаимосвязей юнитов на энергетическом ландшафте в процессе минимизации информационной энтропии и локальной минимизации термодинамической энтропии.*

Выводы. Результаты данных исследований говорят о концептуальных ограничениях моделей ИНС первого и второго типов. Структура внутренней энергии этих систем в состоянии покоя формируется весовыми коэффициентами, которые получены другой системой - "Демоном Максвелла", которая является внешней по отношению к рассматриваемым системам. Эти весовые коэффициенты формируются на основе статистических наблюдений и никак не связаны с внешней энергией (информацией). Эти ограничения не позволяют реализовать процесс самоорганизации системы на основе энергии, как это происходит в реальных физических системах, например, квантовых.

6. Информация

6.1. Эволюция энергетического ландшафта. Структурная и параметрическая редукция.

Определим общую внутреннюю энергию системы E_{total} как совокупность внутренних энергий ее структурных элементов (юнитов) и взаимосвязей между ними

$$E_{total} = \sum_i E_i + \sum_{i < j} E_{ij}$$

(43),

где E_i – внутренняя энергия структурного элемента i , E_{ij} – энергия взаимосвязи между элементами i и j .

Обозначим как N_{true} число структур входной энергии (событий) истинных для заданного макросостояний системы. В данном случае, под истинностью мы будем понимать соответствие структуры входной энергии заданному макросостоянию системы. В современных ИНС под истинными событиями понимаются примеры

обучающей выборки. Тогда, N_{total} – общее число событий, воспринимаемых системой. Мы можем определить весовой коэффициент истинных событий w , которые влияют на энергию системы

$$w = \frac{N_{true}}{N_{total}} \quad (44).$$

Эту зависимость мы обозначим как

$$E_i = E_i(w), \quad E_{ij} = E_{ij}(w) \quad (45).$$

Определим, как будет изменяться E_{total} при увеличении числа истинных событий N_{true} .

$$\frac{dE_{total}}{dN_{true}} = \sum_i \frac{\partial E_i}{\partial w} \cdot \frac{\partial w}{\partial N_{true}} + \sum_{i < j} \frac{\partial E_{ij}}{\partial w} \cdot \frac{\partial w}{\partial N_{true}} \quad (46).$$

Весовой коэффициент w увеличивается при увеличении количества истинных событий N_{true} . Для демонстрации этого процесса найдем его производную по N_{true} .

$$\frac{dw}{dN_{true}} = \frac{d}{dN_{true}} \left(\frac{N_{true}}{N_{total}} \right) \quad (47).$$

Если общее количество событий N_{total} фиксировано (например, ограничено размером обучающей выборки), то производная принимает простой вид

$$\frac{dw}{dN_{true}} = \frac{1}{N_{total}}$$

(48).

Тогда

$$\frac{dE_{total}}{dN_{true}} = \frac{1}{N_{total}} \left(\sum_i \frac{\partial E_i}{\partial w} + \sum_{i<j} \frac{\partial E_{ij}}{\partial w} \right)$$

(49).

При выполнении условий

$$\sum_i \frac{\partial E_i}{\partial w} < 0 \text{ and } \frac{\partial E_{ij}}{\partial w} < 0$$

(50),

функция энергии системы E_{total} также уменьшается:

$$\frac{dE_{total}}{dw} = \sum_i \frac{\partial E_i}{\partial w} + \sum_{i<j} \frac{\partial E_{ij}}{\partial w} < 0$$

(51).

Это условие выполняется только в том случае, если при увеличении статистической значимости структурных элементов или связей, параметры их энергий переходят к более устойчивым, инвариантным значениям. Это значит, что их внутренняя энергия должна иметь структуру, например, в виде дискретных энергетических уровней. Тогда, увеличение весового коэффициента определяет параметрическую общность энергии, т.е. переход на общий, более энергетически

устойчивый уровень. Этот процесс уменьшения энергии можно сравнить с переходом от количественных значений отдельных параметров энергии к более общим и инвариантным качественным значениям параметров.

Таким образом, рост w можно интерпретировать как процесс упорядочивания системы, при котором увеличение статистической значимости "истинных" событий приводит к изменению внутренних параметров так, что система переходит от менее упорядоченного (высокоэнергетического) состояния к более упорядоченному (низкоэнергетическому) состоянию. Математически это отражается в отрицательной производной dE_{total}/dw (51), а физически – в переходе через фазовый переход к состоянию с минимальной энергией.

Аналогичный процесс уменьшения энергии происходит и в квантовых системах, например, при переходе электронов на менее энергетические орбитали.

Таким образом, соотношения (46-50) определяют эволюция энергетического ландшафта системы (ее структуры и энергии) в процессе обучения, а условия (51) определяют взаимозависимость энергии юнитов и их взаимосвязей, т.е. процесс формирования внутренней энергии системы. Эти условия являются основой для построения правил структурной и энергетической (параметрической) редукции системы.

Этап структурной редукции связан с уменьшением количества весовых (статистически значимых) структурных элементов и взаимосвязей между ними. Редукция структуры до статистически значимой непосредственно ведет к снижению общей энергии системы.

Для иллюстрации этого процесса введем порог статистической значимости для весовых коэффициентов юнитов w_i и их взаимосвязей w_{ij} : $w_i \geq \gamma$, $w_{ij} \geq \gamma$, а также оператор редукции структуры R_w , который исключает статистически менее значимые компоненты структуры, что приводит к формированию редуцированной структуры с энергией

$$E_{red} = R_w(E_{total}) < E_{total} \quad (52).$$

Этап параметрической редукции внутренней энергии системы связан с уменьшением энергии юнитов и их взаимосвязей.

Для выполнения этого условия внутренние энергии юнитов E_i и взаимосвязей между ними E_{ij} должны быть взаимозависимы.

Обозначим:

$U^{(0)} = \{u_i^{(0)}\}$ – как множество исходных количественных переменных - измеряемых параметров, характеризующих состояние энергии юнитов и определяющих структуру их энергии;

$C^{(0)} = \{c_{ij}^{(0)}\}$ – как множество исходных вычисляемых параметров (например, на основе выполнения *функции сравнения измеряемых параметров энергии юнитов*), характеризующих состояние энергии взаимосвязи юнитов и определяющих структуру их энергии. В общем случае

$$c_{ij} = \Phi(u_i, u_j) \quad (53),$$

где Φ – функция сравнения, которая задает параметр взаимосвязи c_{ij} как результат сравнения, в данном случае, параметров двух юнитов. Таким образом, уравнение (11) связывает энергию юнитов с энергией из взаимосвязей. Кроме того, эта функция формирует внутренний энергетический ландшафт взаимосвязей юнитов - "каналов" на общем энергетическом ландшафте системы (Рис.1). Это происходит в результате дискретизации уровней энергии при переходе от количественных к качественным значениям параметров сравнения, т.е. при переходе от количественной шкалы измерений к качественной шкале. В общем случае, существует два типа качественных шкал: номинальная и порядковая, и два типа количественных шкал: интервальная и шкала отношений, которые являются

метрическими [25]. Наиболее "сильной" шкалой является шкала отношений, т.к. она формирует абсолютную метрику количественных значений параметров. Т.е. мы можем интерпретировать ее "силу" как наиболее высокий энергетический уровень. В то время, как качественная номинальная шкала только разделяет категории признаков без их количественного измерения или упорядочивания. Это самая "слабая" шкала, с которой мы можем сопоставить самый низкий уровень энергии. Тогда, наиболее инвариантной к количественным значениям параметров и, в то же время, наиболее информативной качественной шкалой для определения энергии взаимосвязи, является порядковая шкала. Метафорически ее энергию можно сравнить с энергией p орбитали атома, которая, благодаря своей направленности и роли в установлении связей, занимает центральное место в процессе взаимосвязи атомов [26].

На каждом структурном уровне $L = 0, 1, 2, \dots, L_{max}$ система характеризуется:

- множеством параметров юнитов $U^{(L)} = \left\{ u_i^{(L)} \right\}$,
- множеством параметров взаимосвязей $C^{(L)} = \left\{ c_{ij}^{(L)} \right\}$.

При этом общая энергия системы E_{total} на уровне L будет иметь вид

$$E_{total}^{(L)} = E_{units}^{(L)}(U^{(L)}) + E_{int}^{(L)}(U^{(L)}, C^{(L)}) \quad (54),$$

где

$$E_{units}^{(L)}(U^{(L)}) = \sum_i E_i^{(L)}(u_i^{(L)}) \quad (55),$$

$$E_{int}^{(L)}(U^{(L)}, C^{(L)}) = \sum_{ij} E_{ij}^{(L)}(c_{ij}^{(L)}) = \sum_{ij} E_{ij}^{(L)}\left(\Phi(u_i^{(L-1)}, u_j^{(L-1)})\right) \quad (56).$$

При переходе с уровня L на уровень $L + 1$ производится редукция параметров – от исходных количественных характеристик к более инвариантным качественным, что приводит к уменьшению энергии

$$E_{total}^{(L+1)} < E_{total}^{(L)} \quad (57).$$

При этом, динамика изменения параметров описывается градиентным спуском по энергетическим ландшафтам отдельных юнитов и их взаимосвязей на каждом уровне в процессе эволюционного формирования общего энергетического ландшафта системы, т.е. в процессе обучения. Для юнитов и их взаимосвязей можно записать следующие итерационные соотношения:

$$U^{(L+1)} = U^{(L)} - \eta^{(L)} \frac{\partial E_{units}^{(L)}}{\partial U^{(L)}} \quad (58),$$

$$C^{(L+1)} = C^{(L)} - \eta^{(L)} \frac{\partial E_{int}^{(L)}}{\partial C^{(L)}} \quad (59),$$

где $\eta^{(L)} > 0$ – шаг перехода между шкалами измерений.

Этот процесс можно сравнить с энергетическими процессами в классической механике, где силы в системе связаны с градиентом потенциальной энергии: $F = - \Delta U$ (где U – потенциальная энергия). Это означает, что если система оказывается в состоянии с более высокой энергией, то действующие силы будут направлены так, чтобы перевести систему в состояние с более низкой энергией. Таким образом, любые флуктуации, повышающие энергию, "отталкиваются" силовыми полями, приводящими систему обратно к минимуму потенциальной энергии.

Условие минимизации энергии на каждом уровне требует, чтобы в предельном случае выполнялись условия равновесия:

$$\frac{\partial E_{total}^{(L)}}{\partial u_i^{(L)}} = \frac{\partial E_{units}^{(L)}}{\partial u_i^{(L)}} + \frac{\partial E_{int}^{(L)}}{\partial u_i^{(L)}} = 0 \quad \forall i$$

(60),

$$\frac{\partial E_{total}^{(L)}}{\partial c_{ij}^{(L)}} = \frac{\partial E_{int}^{(L)}}{\partial c_{ij}^{(L)}} = 0 \quad \forall (i, j)$$

(61).

При этом выполняется взаимозависимость:

$$\frac{\partial^2 E_{int}^{(L)}}{\partial u_i^{(L)} \partial c_{ij}^{(L)}} \neq 0$$

(62),

которая гарантирует, что изменение параметров взаимосвязей $c_{ij}^{(L)}$ влияет на оптимальное распределение $u_i^{(L)}$ и наоборот. В частности, если изменение параметров связей приводит к уменьшению E_{int} , например, в результате установления более энергетически "выгодной" связи, то условие равновесия для юнитов (60) заставляет систему изменить u_i таким образом, чтобы E_{units} также уменьшилась. Такая же взаимозависимость энергий выполняется в моделях физических систем, например, в модели Изинга.

Чтобы упростить анализ на каждом уровне, можно выразить параметры юнитов как функцию от параметров взаимосвязей, то есть $U^{(L)} = U^{(L)}(C^{(L)})$. Тогда, определяется *эффективная энергия*:

$$E_{eff}^{(L)}(C^{(L)}) = E_{units}^{(L)}\left(U^{(L)}(C^{(L)})\right) + E_{int}^{(L)}\left(U^{(L)}(C^{(L)}), C^{(L)}\right) \quad (63).$$

Понятие эффективной энергии позволяет описывать сложные системы, упрощая их анализ, т.е. учитывать микроскопические процессы через их усреднение, чтобы сосредоточиться на макроскопическом поведении системы [27].

Полный дифференциал эффективной энергии по параметрам взаимосвязей имеет вид:

$$\frac{dE_{eff}^{(L)}}{dC^{(L)}} = \frac{\partial E_{int}^{(L)}}{\partial C^{(L)}} + \left[\frac{\partial E_{units}^{(L)}}{\partial U^{(L)}} + \frac{\partial E_{int}^{(L)}}{\partial U^{(L)}} \right] \frac{dU^{(L)}}{dC^{(L)}} \quad (64).$$

В условиях равновесия, согласно условиям минимизации по $U^{(L)}$ (60), в квадратных скобках выражение равно нулю. Тогда:

$$\frac{dE_{eff}^{(L)}}{dC^{(L)}} = \frac{\partial E_{int}^{(L)}}{\partial C^{(L)}} \quad (65).$$

Следовательно, редукция эффективной энергии происходит за счёт изменения параметров взаимосвязей. Это соотношение иллюстрирует, что уменьшение энергии взаимосвязи E_{int} за счет изменения C автоматически ведет к изменению состояния U таким образом, что энергия юнитов E_{units} также снижается до оптимального состояния в результате перехода от количественных к менее энергетическим качественным параметрам.

Тогда

$$\frac{dE_{eff}^{(L)}}{dC^{(L)}} < 0$$

(66).

Таким образом, на этапе параметрической редукции $R_{u,c}$ также происходит уменьшение общей энергии системы

$$E_{red} = R_{u,c}(E_{total}) < E_{total}$$

(67).

6.2. Структурно-параметрическая редукция.

Одним из основных этапов редукции, которая формирует структуру-аттрактор (концепт), представляющую собой структуру энергетического ландшафта системы в состоянии покоя, является *структурно-параметрическая редукция*.

Эта редукция позволяет выделять подструктуры с неизменными параметрами взаимосвязи между элементами (юнитами) и упрощение ее в результате сокращения элементов, которые не оказывают существенное влияние на параметрическую устойчивость подструктуры в целом. Эти элементы являются наименее информативными, что позволяет сжать подструктуру до двух элементов: двух экстремальных или критических элементов - критических структурных точек: начальной и конечной точек подструктуры, взаимосвязь между которыми будет иметь те же общие параметры, что и для всех элементов рассматриваемой подструктуры. Мы можем провести аналогию этого процесса с процессом сжатия информации в информационных системах, а сами критические точки мы можем сравнить с точками бифуркации в теории диссипативных или динамических систем.

Рассмотрим данный вид редукции. Пусть существует последовательно упорядоченная структура S , состоящая из элементов s_i , где каждый элемент характеризуется множеством однотипных параметров u_i .

$$S = (s_1, s_2, \dots, s_n), \quad s_i = \{u_{i1}, u_{i2}, \dots, u_{im}\} \quad (68).$$

В этой последовательности структурных элементов выбираем первый элемент в качестве начальной критической точки устойчивой подструктуры, которую мы пытаемся определить. Нам необходимо найти конечную критическую точку и редуцировать найденную подструктуру. Введем операцию сравнения Φ однотипных параметров u_k двух соседних структурных элементов s_i и s_{i+1} , которая формирует их взаимосвязь $c_{i,i+1}$ аналогично формуле (53).

$$c_{i,i+1} = \Phi(u_{i,k}, u_{i+1,k}) \quad (69).$$

Если параметры являются количественными, то значение их разности отображается на интервальную шкалу функцией φ .

$$\Phi(u_{i,k}, u_{i+1,k}) = u_{i,k} - u_{i+1,k} = \varphi(\Delta u_{i,i+1}^k) \quad (70).$$

Если для следующей структурной пары элементов $u_{i+1,k}$ и $u_{i+2,k}$, упорядоченной по индексу i , значение разности этих параметров будет отличаться от первой пары, то для приведения параметров взаимосвязей $c_{i,i+1}$ и $c_{i+1,i+2}$ к единому значению u_k , используем первую производную

$$\frac{du_k}{ds_i} \approx u_{i,k} - u_{i+1,k} = u_{i+1,k} - u_{i+2,k} \quad (71).$$

В данном случае, для поиска инвариантного значения для обеих взаимосвязей нас будет интересовать даже не само значение производной, а ее градиент, т.е. знак этой разности.

$$\text{sign}\left(\frac{du_k}{ds_i}\right) = \begin{cases} +1, & \text{parameter increase} \\ 0, & \text{parameter constancy} \\ -1, & \text{parameter decrease} \end{cases}$$

(72),

Тогда, в случае совпадения знаков градиента

$$c'_{i,i+1} = c'_{i+1,i+2} = \sigma\left(\text{sign}\left(\frac{du_k}{ds_i}\right)\right)$$

(73),

где σ – функция отображения градиента (производного качественного параметра) на качественную шкалу порядка.

Для структуры из n элементов можно рассмотреть вектор качественных параметров взаимосвязей для всех пар соседних элементов

$$C'_S = (c'_{1,2}, c'_{2,3}, \dots, c'_{n-1,n})$$

(74).

Если все элементы этого вектора равны, но следующий элемент $c'_{n,n+1}$ имеет другое значение $c'_{n-1,n} \neq c'_{n,n+1}$, то структурный элемент s_n является конечной критической точкой подструктуры устойчивой по параметру u_k , редуцированному до качественного значения u'_k . В этом случае, мы можем осуществить структурно-параметрическую редукцию R_{sp}

$$S' = R_{sp}(S) = \left[s_1(u'_{1,k}) \right] c'_{1,n} \left[s_n(u'_{n,k}) \right]$$

(75).

Действительно, если параметры взаимосвязей $c'_{i,i+1}$ структурных элементов s_i и s_{i+1} равны, то в соответствии с условием взаимозависимости энергии юнитов и энергии их взаимосвязи (63) будут равны и редуцированные значения параметров структурных элементов (параметры энергии юнитов) $u'_{i,k}$ и $u'_{i+1,k}$. Тогда, мы можем упростить структуру S до структуры S' (75) по параметру u'_k путем удаления повторяющихся структурных элементов, т.е. осуществить сжатие структуры, что будет соответствовать выполнению условий уменьшения эффективной энергии (65, 66).

Однако, при больших значениях разности (или расстояния) количественных значений параметров $u_{l,k}$ и $u_{n,k}$ может быть потеряна существенная информация о величине их изменения относительно друг друга. Для устранения этой проблемы необходимо ввести качественные дискретные пороги количественного изменения данного параметра в соответствии с заданной шкалой или системой координат.

Например, при рассмотрении такого параметра как вектор ориентации отрезка или его местоположение в прямоугольной системе координат, мы можем сегментировать эту систему координат на множество обобщенных (качественных) подструктур: полуплоскости, квадранты или сегменты квадрантов. В данном случае, эта сегментация и будет определять пороги значений параметров ориентации или местоположения, т.е. определять пределы изменения количественных параметров и критические структурных точки как точки преодоления этих пороговых значений.

Тогда, для уравнения (75) нужно записать ограничения для параметров $u_{l,k}$ и $u_{n,k}$

$$u_{1,k} \geq Tr_j, \quad u_{n,k} < Tr_{j+1} \quad (76),$$

где Tr_j и Tr_{j+1} – пороги значений параметров, определяющие их принадлежность к одному сегменту шкалы или системы координат. В этом случае, если $u_{n,k} \geq Tr_{j+1}$, то

в структуре S определяется критическая точка даже при неизменном значении параметров $c'_{i,i+1}$. Переходы между сегментами шкал или систем координат разной степени общности, например, при увеличении общности сегментов прямоугольной системы координат от сегментов квадранта к квадрантам и полуплоскостям, также определяют шаги уменьшения внутренней энергии системы – спуска по поверхности ее энергетического ландшафта.

Определение значений порогов, как и сама сегментация шкал или систем координат, должны осуществляться в процессе адаптации (обучения) системы детектирования (измерения) параметров воспринимаемой энергии.

Очевидно, что используя градиент второй производной, мы переходим к еще одной качественной шкале, анализируя ускорение или замедление изменений параметров.

Очевидно, что качественных (производных) константных параметров, которые определяют устойчивые подструктуры, может быть множество. В этом случае формируется множество различных параметрических взаимосвязей между последовательностью структурных элементов. Тогда, установление (выбор) главных параметров для структурно-параметрической редукции осуществляется на основе весовых коэффициентов w , сформированных в процессе обучения (43-51), т.е. в процессе структурной редукции R_w .

Таким образом, должна выполняться следующая последовательность редукций внутренней энергии системы: параметрическая редукция $R_{u,c}$, структурно-параметрическая редукция R_{sp} , структурная редукция R_w .

$$E_{red} = R_w(R_{sp}(R_{u,c}(E_{total}^{(0)}))) < E_{total}^{(0)} \quad (77),$$

где $E_{total}^{(0)}$ – общая энергия системы на начальном структурном уровне.

При этом, переход от количественных к качественным параметрам, шкалам (системам координат) и их сегментам определяют шаги градиентного спуска по поверхности энергетического ландшафта системы. Усиление этих шкал возможно при рассмотрении первоначальных количественных значений разности параметров энергии юнитов (структурных элементов) не в интервальной шкале, а в шкале отношений, когда параметры начальной критической точки определяют абсолютное значение - центр системы координат и осуществляется сравнение параметров выбранной структурной точки с этим центром.

Таким образом, множество структурных элементов (юнитов) их взаимосвязей, параметров, систем координат, шкал и порогов их сегментации, критических структурных точек (точек бифуркации) формируют фазовое пространство рассматриваемой энергетической системы. Редуцированная структура юнитов и их взаимосвязей отображает эволюцию энергетического ландшафта системы в процессе обучения и представляет собой энергетический и структурный аттрактор – концепт для определенного множества (класса) структур измеряемых параметров входной энергии. Тогда можно сделать вывод, что уравнение (77) отображает обобщенное представление энергетической функции самоорганизующейся и эволюционирующей системы в виде функционала Ляпунова.

6.3. Самоорганизация и эволюция системы на основе энергии.

Рассмотренные этапы редукции отображают процесс формирования *устойчивой энергетической структуры – аттрактора*, которая соответствует наиболее общей и инвариантной структуре определенного класса входной информации.

Таким образом, полученная редуцированная структура – аттрактор является внутренней моделью определенного класса пространственно-временной

структуры измеряемых параметров входной энергии. Она определяет множество доступных микросостояний системы, которые соответствуют заданному макросостоянию, определяемому реакцией системы. Назовем эту внутреннюю модель - концептом данного класса.

Аналогичная схема реализуется в квантовых системах с использованием семейства гамильтонианов

$$H(\lambda) = H_{units} + \lambda V \quad (78).$$

Здесь H_{units} представляет собой часть гамильтониана, описывающую внутренние (или "единичные") свойства системы без учета взаимодействия, а λV – взаимодействие, сила которого регулируется параметром λ . Согласно теореме Хеллмана-Фейгенбаума [28], если $\Psi(\lambda)$ – нормированное собственное состояние $H(\lambda)$ с энергией $E(\lambda)$, то

$$\frac{\partial E}{\partial \lambda} = \langle \Psi(\lambda) | \frac{\partial H}{\partial \lambda} | \Psi(\lambda) \rangle \quad (79).$$

Так как зависимость гамильтониана от λ задается через λV , то имеем $\frac{\partial H}{\partial \lambda} = V$ и, соответственно,

$$\frac{\partial E}{\partial \lambda} = \langle \Psi(\lambda) | V | \Psi(\lambda) \rangle \quad (80).$$

Если оператор V описывает притягательное (энергосжимающее) взаимодействие, то его матожидание $\langle \Psi(\lambda) | V | \Psi(\lambda) \rangle$ имеет отрицательное значение.

Это означает, что при увеличении λ (то есть при усилении взаимодействия) энергия системы $E(\lambda)$ уменьшается (становится более отрицательной).

Важно отметить, что снижение энергии не происходит только за счет вкладов от λV . Волновая функция $\Psi(\lambda)$ при этом также изменяется, чтобы обеспечить оптимизацию (минимизацию) общей энергии. В результате адаптация $\Psi(\lambda)$ приводит к тому, что вклад от H_{units} также уменьшается. Иными словами, система настраивает свое внутреннее состояние таким образом, чтобы совместно с усилением притягательного взаимодействия достигать более низкого энергетического состояния.

Следовательно, в квантовых системах наблюдается та же взаимозависимость: установление выгодных (энергоснижающих) взаимосвязей приводит к снижению как энергии их взаимосвязей, так и энергии отдельных юнитов.

В статистической физике вероятность того, что система находится в состоянии с энергией E , определяется распределением Больцмана (20). Это означает, что при заданной температуре состояния с меньшей энергией значительно более вероятны. В термодинамическом равновесии система практически всегда оказывается в конфигурации, соответствующей минимуму свободной энергии, что обеспечивает максимальную статистическую вероятность и стабильность.

Таким образом, в рассматриваемой схеме уравнение редукции энергии системы (77), наряду с уравнениями эффективной энергии (63, 64) и функцией энергии системы E_{total} (51), не только иллюстрируют, как происходит уменьшение энергии системы, но и фактически определяют динамику процесса ее минимизации. Все эти уравнения выступают в роли функционала Ляпунова: все функции монотонно убывают при эволюции системы, направляя её к устойчивому энергетически оптимальному состоянию. В отличие от энергетической функции в модели Хопфилда, которая используется лишь для демонстрации сходимости

динамики, эти энергетические функции непосредственно управляют процессом редукции за счёт следующих особенностей:

- взаимозависимость параметров: поскольку внутренние энергии юнитов и энергия их взаимосвязей зависят друг от друга (условия (53, 60–62)), изменение параметров взаимосвязей C вынуждает систему корректировать параметры U так, чтобы минимизировать E_{units} . Это обеспечивает согласованное снижение полной энергии E_{total} (51);

- прямое управление динамикой на основе параметрической редукции: благодаря условиям равновесия (60-62), минимизация $E_{eff}(C)$ (64) становится не просто демонстрацией, а механизмом, по которому система "саморегулируется". Изменения во взаимодействиях приводят к оптимизации как энергий структурных единиц, так и их взаимосвязей;

- прямое управление динамикой на основе структурно-параметрической редукции: при выполнении условий (71-73, 76) система эволюционирует к устойчивым поструктурам (75) с минимальной энергией:

- прямое управление динамикой на основе структурной редукции: при выполнении условий (46-50) система уменьшает свою общую энергию (51) в результате оптимизации своей структуры (52).

Таким образом, уравнение (77) описывает общую концепцию самоорганизующейся и эволюционирующей энергетической системы, где упорядоченность системы на основе минимизации ее внутренней энергии определяет эволюцию к устойчивому состоянию - аттрактору, аналогично тому, как это происходит в физических или квантовых системах.

В этой концепции этапы (шаги) редукции ограничивают количество микросостояний системы для заданного макросостояния. Таким образом, система может "абстрагироваться" от микросостояний (энергетических элементов и связей), которые оказывают несущественное или неустойчивое влияние

(например, помехи) на достижение заданного макросостояния, т.е. с точки зрения термодинамики, уменьшается доступное фазовое пространство системы.

Однако это возможно только для единичного (не обобщенного) ландшафта, т.е. с одним макросостоянием. Тогда, подобные эволюционные системы (например, эволюционные ИНС), должны строить обобщенные энергетические ландшафты в виде взаимосвязи одиночных энергетических ландшафтов, как это и происходит в сложных физических системах, а не на основе их объединения и усреднения, как это происходит в современных коннекционистских ИНС. Такая концепция взаимосвязи одиночных нейронов (например, на основе механизма конкуренции), которые реагируют на целостные образы, хорошо известна в нейронауках, как теория одиночных нейронов или "бабушкиных клеток" [29-31].

Рассмотрим термодинамику процесса структурной редукции, в которой центральную роль играют значения весовых коэффициентов w . Если S – энтропия Больцмана, а Ω - число доступных микросостояний, то весовые коэффициенты w ограничивают их количество

$$\Omega(w) = \Omega_0 w^\gamma, \gamma > 0 \quad (81),$$

где $\Omega(w)$ – количество доступных микросостояний для данного макросостояния системы при параметре w ;

Ω_0 – исходное (общее) количество микросостояний, которое система могла бы иметь при $w = 1$;

w^γ - нелинейная зависимость числа микросостояний от w , где $\gamma > 0$ – положительный параметр, который регулирует темп уменьшения числа микросостояний с ростом w . В реальных физических системах количество доступных микросостояний обычно не изменяется линейно, в зависимости от изменения структуры.

Тогда, энтропия системы будет иметь следующий вид

$$S(w) = k_B \ln \Omega_0 + k_B \gamma \ln w \quad (82).$$

Таким образом, увеличение w ведет к локальному снижению термодинамической энтропии для заданных макросостояний системы, что также является подтверждением наших предыдущих выводов.

Теперь мы можем рассмотреть E_{total} как

$$E_{total} = E_0 - TS(w) \quad (83),$$

где E_0 – начальная энергия системы (до обучения) или ее энергия при максимальной энтропии ($w = 1$), когда все возможные микросостояния доступны;

T – температура (или аналогичный параметр, характеризующий степень случайных флуктуаций).

Тогда

$$E_{total} = E_0 - k_B \ln \Omega_0 - k_B \gamma \ln w \quad (84)$$

и

$$\frac{dE_{total}}{dw} = - \frac{k_B T \gamma}{w} < 0, \quad when \ w > 0 \quad (85).$$

То есть, энергия действительно убывает с ростом w в результате уменьшения фазового пространства доступных микросостояний.

Однако, ограничение числа микросостояний на этапе структурной редукции (52) не означает простого "выкидывания" элементов или связей из структуры системы. В физических системах такие ограничения должны приводить к перестройке внутренней структуры системы. Эта перестройка важна, так как

система должна оставаться целостной и сохранять свой макроскопический порядок. Т.е. система должна *эволюционировать* в результате реорганизации своих структурных элементов E_i и связей E_{ij} так, чтобы новые микросостояния, которые будут доступны после редукции, поддерживали макроскопическое состояние системы. При уменьшении доступных микросостояний система должна оптимизировать связи между элементами, чтобы снизить энтропию и сохранить целостную структуру концепта.

Таким образом, для самоорганизации и эволюции энергетической системы, должны выполняться следующие правила:

- *правила минимизации энергии*, когда система должна стремиться к состоянию, где её внутренняя энергия элементов и их взаимодействий, минимальна при ограничениях, определяемых параметрами w и градиентным спуском по их энергетическим ландшафтам;
- *правила оптимизации структуры*, когда система должна не только минимизировать свою энергию, но и оптимизировать свою структуру, не только оставляя те связи и элементы, которые важны для поддержания макросостояния, но и создавая новые более устойчивые связи между элементами для формирования концепта.

При этом должны соблюдаться несколько принципов:

- *принцип локальной минимизации энергии*, когда каждый элемент системы должен стремиться минимизировать свою собственную энергию при учёте окружающих его связей. Это должно происходить в результате изменение взаимодействий с соседними элементами, что приведет к переходу в более устойчивое состояние. Этот принцип можно выразить как

$$\frac{dE_i}{dx_i} = 0$$

(86),

где x_i – параметры, описывающие положение или состояние элемента;

- *принцип минимизации энтропии*, когда, для того чтобы система оставалась структурированной, она должна поддерживать максимально упорядоченное состояние при минимальной энтропии. Этот принцип отражает стремление системы уменьшить неустойчивые состояния (например, неважные связи или элементы), которые могут вызвать флуктуации или помехи. В случае рассматриваемой концепции с весовыми коэффициентами w , которые уменьшают число доступных микросостояний, принцип минимизации энтропии может быть представлен следующим образом:

$$\frac{dS}{dw} < 0$$

(87);

- *принцип максимизации связности (или топологической устойчивости)*, когда, для того чтобы структура системы оставалась целостной и устойчивой, элементы должны сохранить общую топологию сети, даже если исключаются некоторые связи или параметры. Это означает, что система не должна "разваливаться" или терять свои ключевые функции или качества. Математически это может быть связано с концепцией топологической устойчивости, где система должна избегать ситуаций, в которых её структура становится разрозненной. Этот принцип можно реализовать с использованием графов или сетей.

Таким образом, мы видим совершенно иную роль весовых коэффициентов в энергетической модели ИНС по сравнению с классической моделью, где весовые коэффициенты формируют глубину потенциальных "ям" юнитов либо взаимосвязей между ними в состоянии покоя. В рассматриваемой модели формирование внутренней энергии юнитов и взаимосвязей между ними должно осуществляться на основе внешней энергии, а не на основе "псевдоэнергии" (например, энергии ошибки), как это происходит в современных

коннекционистских ИНС. При этом, построение структуры этой внутренней энергии и ее минимизация осуществляется на основе правил редукции под управлением энергетических функций, моделирующих условия энергетического взаимодействия (взаимозависимости), что является моделью, вдохновленной энергетическими процессами, происходящими в физических системах.

Общие выводы относительно процессов самоорганизации и эволюции систем.

1. Самоорганизация (как формирование энергетического ландшафта), а также эволюция (как изменение энергетического ландшафта) любой энергетической системы возможна только в результате преобразования внешней энергии во внутреннюю энергию системы и установления энергетических взаимосвязей между внутренними энергиями отдельных юнитов. Если юниты и взаимосвязи между ними не обладают взаимосвязанной внутренней энергетической структурой, то очевидно, что такая самоорганизация на основе минимизации внутренней энергии системы невозможна.

2. Для любой энергетической системы, которая обладает общей структурой внутренней энергии юнитов и их взаимосвязей существует энергетическая функция, которая обуславливает минимизацию энергии.

Основой этого утверждения может служить тот факт, что если рассматривать системы, для которых можно построить гладкий функционал, описывающий энергию, и если динамика этой системы имеет градиентную природу или может быть сведена к процессу, сопровождаемому монотонным уменьшением этой энергии, то для таких систем всегда можно построить функцию Ляпунова и интерпретировать ее как энергофункцию.

Мы можем не знать эту функцию, если нам не известна структура энергии, как, например, в диссипативных системах Пригожина. Тогда, мы можем применить другой механизм анализа процессов минимизации энергии. Однако, если мы строим детерминированную систему (например, искусственную нейронную сеть), которая не основана лишь на вероятностном или статистическом

подходе, то для организации процессов самоструктуризации нам необходима общая структура внутренней энергии.

3. Если говорить об альтернативных подходах к процессам "самоорганизации", например в ИНС через адаптацию весовых коэффициентов, то это процесс "самоструктурирования" носит условный характер. Для его реализации необходимо участие *внешней системы* ("Демона Максвелла") в том или ином виде. В этом случае именно оптимизированные параметры внешней "псевдоэнергии" (например, энергии ошибки) формируют внутреннюю структуру энергии юнитов (нейронов), как это происходит в классифицирующих ИНС, либо энергии взаимосвязи между юнитами, как это происходит в генеративных ИНС.

4. Принцип минимизации внутренней энергии лежит в основе самоорганизации и эволюции не только любой физической системы, но и любой информационной системы, если под "внутренней энергией" понимать любую величину, которая характеризует состояние системы и подлежит оптимизации. Всё сводится к тому, что самоорганизация — это процесс оптимизации, а выбранная интерпретация "энергии" определяет, каким образом этот процесс описывается в конкретном контексте.

5. Основным проявлением ограничений для процесса самоорганизации в ИНС является наличие "Демона Максвелла", который необходим для минимизации внешней "псевдоэнергии". Структура внутренней энергии системы, формируемая под воздействием параметров "псевдоэнергии" не в полной мере отражает возможные измеряемые параметры входной энергии, т.е. информацию на входе системы. Это ведет к тому, что параметры "псевдоэнергии" влияют либо на формирование внутренней энергии юнитов в ИНС первого типа, либо на формирование внутренней энергии взаимосвязей между юнитами в ИНС второго типа. Тогда как, при самоорганизации системы на основе внутренней энергии, которая формируется на основе входной информации, мы можем отказаться от "Демона" и эволюционно формировать энергетический ландшафт в состоянии

покоя. Для такого процесса не важно, как физически представлена внешняя энергия, но важно какая информация может быть получена системой из этой энергии для формирования ее внутренней энергии.

6.4. Внутренняя модель внешнего мира. Информация.

Можно говорить о том, что внутренняя структура самоорганизующейся энергетической системы представляет собой *модель внешнего мира системы M* , которая сформированная в процессе: восприятия структурированной внешней энергии (детектирования и измерения ее параметров) и ее преобразования во внутреннюю энергию, построения (самоорганизации) первоначальной структуры энергетического ландшафта системы в состоянии покоя, а также ее редукции (эволюции) в процессе обучения.

$$M = R(S_{int}(P(E_{int}))) = Con \quad (88),$$

$$P(E_{int}) = D(S_{ext}(E_{ext})) \quad (89),$$

где R – обобщенный оператор редукции, $S_{int}(P(E_{int}))$ – структура множества параметров P внутренней энергии системы E_{int} , D – обобщенный оператор детектирования (измерения) структурированных S_{ext} (имеющих пространственно-временную структуру) параметров внешней энергии E_{ext} .

Модель M можно представлять в виде гиперграфа, а ее эволюцию – в виде гиперсети [32]. Использование подобного представления также позволяет формально доказать единственность концепта Con - аттрактора для заданного класса входных информационных векторов, его параметрическую и структурную устойчивость, а также сходимости исходной структуры к аттрактору за конечное число шагов редукции.

Если мы рассмотрим процесс *интерпретации* этой моделью входной энергии и формирование выходной энергии или информации на выходе системы, то мы можем его сопоставить с процессом инференса в ИНС. В процессе интерпретации система воспринимает входную энергию и строит ее текущую модель $M'(t)$, которая редуцируется и сравнивается с концептом (запомненной моделью) M .

$$M'(t) = S_{int}(P(E_{int}(t))) \quad (90).$$

Представим данные модели в виде гиперграфов $G' = R(M'(t))$ и $G = M$. Если G' изоморфен G

$$f: G' \rightarrow G \quad (91),$$

где f – функция взаимно-однозначного отображения (биекция), то $R(M'(t)) = M$ и $M' \in M$.

Если G является подграфом G' (субграфовый изоморфизм)

$$f': G \rightarrow G' \quad (92),$$

где f' – инъективная функция, то также выполняется условие интерпретации $M' \in M$, или

$$M = Int(M') \quad (93),$$

где Int – обобщенный оператор интерпретации модели M' моделью M .

Если условия (91) или (92) не выполняются, то M' не может быть интерпретирована моделью M , т.е. структура и значения параметров входной

энергии (входная информация) не совпадают с запомненной энергией (энергетическим ландшафтом системы в состоянии покоя).

В результате интерпретации модели M' происходит формирование реакции - выходной информации I или выходной энергии системы в виде структуры параметров излучаемой энергии $P(E_{out})$, которые являются отображением соответствующих параметров внутренней энергии системы. Формирование этой структуры параметров на выходе системы осуществляется активационной функцией (или функцией чтения) F

$$F(M) = S_{out}(P(E_{out})) = I$$

(94),

где S_{out} – структура (например, пространственно-временная) параметров внешней энергии $P(E_{out})$. Эти выводы согласуются с интерпретацией (33) свободной энергии Гиббса как структурированной части энергии на выходе системы (энтальпии). Это также подтверждается согласованным поведением информационной и термодинамической энтропии как на выходе отдельной модели нейрона, так и ИНС в целом.

Реакция I системы может быть константной (неизменной) для определенного класса входной энергии при $F(M)$, т.е. зависящей от структуры концепта, или динамической при $F(M')$, если реакция системы зависит от текущего состояния ее энергетического ландшафта. Например, в простейшем случае константная реакция системы может отражать сумму структурных элементов концепта. Такая реакция позволяет реализовать процесс конкуренции типа WTA (Winner Take All) между реакциями соседних концептов, представленных, например, искусственными или биологическими нейронами.

Таким образом, на основании приведенных исследований можно дать обобщенное формальное определение информации.

Определение. Структура субъективно воспринимаемых и измеряемых системой параметров внешней энергии для формирования структуры взаимосвязанных параметров ее внутренней энергии – модели внешнего мира, или для ее интерпретации уже существующей моделью, является информацией для данной системы.

Если система воспринимает внешнюю энергию как условную энергию в виде структуры элементов внутреннего алфавита системы (например, вектора входных сигналов), как это происходит в современных информационных системах и ИНС, то модель внешнего мира такой системы, как и процесс интерпретации вектора сигналов, может строиться только на основе статистической взаимосвязи элементов этого алфавита, что приводит к пониманию информации с точки зрения информационной энтропии Шеннона.

Данное определение информации позволяет рассмотреть *информационное разнообразие моделей* $Div(I)$ как разность моделей M' и M при выполнении условий (91) или (92)

$$Div(I) = M' - M \quad (95).$$

Метрикой этого разнообразия может служить, например, редакционное расстояние между графами.

Энергетическая система, которая описывается уравнениями (88, 89 и 94) может классифицироваться как *система с эволюционной структурой энергетического ландшафта, которая самоорганизуется на основе информации*.

Внутренняя энергия в такой системе минимизируется в результате "сжатия" структуры энергетического ландшафта. Это говорит о том, что реакция системы (94) может рассматриваться как кодирование информации о состоянии модели внешнего мира. В сложных информационно-энергетических системах, как например мозг, необходимость такого кодирования обусловлена процессами

дальнейшей обработки информации взаимосвязанными моделями внешнего и внутреннего мира. В такой интерпретации, о рассмотренной системе корректнее говорить не как о "полной" модели внешнего мира, а лишь как о ее функциональном фрагменте.

Кроме того, такая "кодированная" информация на выходе рассмотренной системы будет восприниматься только той системой, которая имеет специфические сенсоры (детекторы), например, как это происходит в процессе обмена информацией между биологическими нейронами. Это снижает универсальность обмена информацией между автономными системами. Для реализации универсального процесса обмена информацией между автономными системами с аналогичными моделями внешнего мира, например между людьми, необходима система "декодирования" информации, т.е. ее дивергенции, которую условно мы можем сравнить с генеративной языковой моделью ИНС. Предположим, человек увидел кота. В цепочке последовательных возбуждений групп нейронов его мозга, произошло возбуждение нейрона целостного образа класса "Кот". Но для передачи этой информации другому человеку необходимо "декодировать" эту реакцию в виде информационной последовательности, например, в виде последовательности звуков или движений руки при написании слова "Кот".

Заметим, что множество видов внутренней энергии в системе зависит от количества и типов сенсоров (детекторов) внешней энергии. Таким образом, именно сенсорные системы образуют мультимодальность внутренней энергии и многообразие фазового пространства энергетической системы. Для искусственных информационно-энергетических систем интерпретация внешней и внутренней энергии зависит от решаемых этими системами задач. Этот взгляд отличается от представления мультимодальности в современных ИНС как различного вида информации: визуальной, звуковой или текстовой. В таких ИНС подобную мультимодальность корректнее было бы назвать мультиагентностью.

