

Кластерный принцип формирования доменов в задачах бинарной минимизации

Литинский Л.Б.

НИИ системных исследований РАН, Москва

litin@mail.ru

Аннотация. При минимизации квадратичного функционала, зависящего от большого числа бинарных переменных, имеет смысл агрегировать переменные, объединяя независимые переменные в более крупные блоки – *домены*. Этот прием позволяет находить более глубокие локальные минимумы. В настоящей работе обосновывается метод формирования доменов, основанный на кластеризации матрицы связей. Результаты компьютерных экспериментов свидетельствуют о том, что кластерные домены позволяют получить более глубокие локальные минимумы, чем домены, сформированные случайным образом.

1. Введение

Обсуждается задача минимизации квадратичного функционала от N бинарных переменных:

$$E(\mathbf{s}) = -(\mathbf{J}\mathbf{s}, \mathbf{s}) = - \sum_{i,j=1}^N J_{ij} s_i s_j \longrightarrow \min_{\mathbf{s}} \text{ где } s_i = \{\pm 1\}. \quad (1)$$

Без ограничения общности матрицу связей $\mathbf{J} = (J_{ij})_1^N$ можно считать симметричной, с нулевыми элементами на главной диагонали: $J_{ij} = J_{ji}$, $J_{ii} = 0$. Используя физическую терминологию [1] бинарные переменные $s_i = \{\pm 1\}$ будем называть *спинами*, N -мерные векторы $\mathbf{s} = (s_1, s_2, \dots, s_N)$ – *конфигурационными векторами* или просто *конфигурациями*, а минимизируемую характеристику $E(\mathbf{s})$ – *энергией* состояния \mathbf{s} .

Общепотребительная процедура локальной минимизации функционала (1) состоит в следующем: стартовав с какой-то начальной конфигурации, случайным образом перебирают спины, и каждому спину присваивают значение, равное знаку действующего на него локального поля

$$h_i(t) = \sum_{j=1}^N J_{ij} s_j(t). \quad (2)$$

Иными словами, если текущее состояние спина $s_i(t)$ совпадает со знаком $h_i(t)$ – если спин *удовлетворен*: $s_i(t)h_i(t) \geq 0$, – то его значение остается неизменным: $s_i(t+1) = s_i(t)$; если же $s_i(t)$ и $h_i(t)$ противоположных знаков – если спин

неудовлетворен: $s_i(t)h_i(t) < 0$, – то в следующий момент времени спин переворачивают: $s_i(t+1) = -s_i(t)$:

$$s_i(t+1) = \begin{cases} s_i(t), & \text{если } s_i(t)h_i(t) \geq 0; \\ -s_i(t), & \text{если } s_i(t)h_i(t) < 0. \end{cases} \quad (3)$$

Легко видеть, что переворот неудовлетворенного спина понижает энергию состояния: $E(\mathbf{s}(t+1)) = E(\mathbf{s}(t)) - 4 |h_i(t)|$. Если позволить системе эволюционировать достаточно долго, рано или поздно она окажется в состоянии, являющемся локальным минимумом по энергии. В этом состоянии все спины будут удовлетворены и эволюция системы прекратится. Данный динамический сценарий будем называть *случайной динамикой*; в теории нейронных сетей его называют асинхронной динамикой [2].

Другой широко известный тип динамики – *синхронная динамика* [2], когда одновременно переворачивают *все неудовлетворенные* спины. Легко показать, что одновременный переворот сразу нескольких неудовлетворенных спинов не гарантирует, что после переворота они станут удовлетворенными. Синхронную динамику редко используют для минимизации функционала (1): для нее невозможно гарантировать монотонность убывания функционала $E(\mathbf{s})$. Кроме того, для синхронной динамики характерно наличие предельных циклов длины 2, когда система “заиклиивается” между двумя состояниями, последовательно перескакивая из одного в другое и обратно. Предельные циклы не имеют никакого отношения к локальным минимумам и только уменьшают их области притяжения.

В работах [3],[4] предложено обобщение случайной динамики – *доменная динамика*. Суть обобщения состоит в том, что спины случайным образом объединяют в более крупные блоки, или (как назвали их авторы) – в *домены*. Затем, при определенных условиях, весь домен переворачивают целиком: все спины, принадлежащие к одному домену, одновременно меняют свои состояния на противоположные. Доменная динамика напоминает синхронную динамику, однако отличается от нее условиями, при которых домен переворачивают. Эти условия естественным образом обобщают случайную динамику на случай агрегированных переменных. Доменная динамика гарантирует монотонное понижение энергии $E(\mathbf{s})$ и не приводит к предельным циклам. Кроме того, доменная динамика в k^2 раз быстрее случайной динамики, где k – средняя длина домена (число объединенных в один блок спинов).

В вычислительных экспериментах со случайными Хеббовскими матрицами¹ было установлено [5], что доменная динамика позволяет отыскать в среднем более глубокие локальные минимумы, чем случайная динамика. Однако остается открытым вопрос – каким образом следует формировать домены? Должны ли домены быть случайными, или имеет смысл формировать их как-то

¹ Хеббовскими называются корреляционные матрицы, элементы которых суть скалярные произведения векторов с бинарными координатами ± 1 ; эти матрицы играют большую роль в теории нейронных сетей [2].

иначе? Зависит ли результат минимизации функционала (1) от способа формирования доменов? Настоящая работа посвящена обсуждению именно этих вопросов.

В следующем разделе дается формальное определение доменной динамики. В третьем разделе обосновывается метод формирования доменов, основанный на идее кластеризации матрицы связей \mathbf{J} . Четвертый раздел посвящен изложению результатов компьютерного моделирования. Несколько общих замечаний вынесено в заключительный раздел, а технические доказательства – в Приложение.

2. Доменная динамика

Для простоты изложения все определения будут даны для случая, когда в первый домен объединены k первых спинов, во второй домен – k следующих спинов, и так далее, в последний, n -й домен, объединены k последних спинов:

$$\mathbf{s} = (\underbrace{s_1, \dots, s_k}_{1\text{-й домен}}, \underbrace{s_{k+1}, \dots, s_{2k}}_{2\text{-й домен}}, \dots, \underbrace{s_{(n-1)k+1}, \dots, s_N}_{n\text{-й домен}}).$$

Здесь $N = kn$, \mathbf{s} – произвольная конфигурация.

Суммарное воздействие на первый домен со стороны всех остальных доменов есть суперпозиция воздействий, которые спины этих доменов оказывают на спины первого домена. Тогда *доменное локальное поле*, действующее на i -й спин первого домена, равно

$$h_i^{(d)}(t) = \sum_{j=k+1}^N J_{ij} s_j(t) = h_i(t) - \sum_{j=1}^k J_{ij} s_j(t), \quad \forall i \leq k, \quad (4)$$

где $h_i(t)$ есть определенное выше локальное поле (2). Иными словами, доменное локальное поле $h_i^{(d)}(t)$ получается из локального поля $h_i(t)$ устранением влияния тех спинов, которые принадлежат вместе с i -м спином к одному и тому же домену. Тогда энергия взаимодействия первого домена с остальными доменами равна

$$E_1(t) = -F_1(t) = -\sum_{i=1}^k s_i(t) h_i^{(d)}(t).$$

Аналогично определяются доменное локальное поле, действующее на спины l -го домена,

$$h_i^{(d)}(t) = h_i(t) - \sum_{j=(l-1)k+1}^{lk} J_{ij} s_j(t), \quad \forall (l-1)k+1 \leq i \leq lk, \quad l \in [2, n],$$

и энергия взаимодействия l -го домена с остальными доменами:

$$E_l(t) = -F_l(t) = - \sum_{i=(l-1)k+1}^{lk} s_i(t)h_i^{(d)}(t), \quad l = 2, \dots, n.$$

Доменная энергия состояния $\mathbf{s}(t)$ есть

$$E^{(d)}(t) = \sum_{l=1}^n E_l(t) = - \sum_{l=1}^n F_l(t). \quad (5)$$

Минимизируемая характеристика $E(t)$ (1) отличается от доменной энергии (5) на сумму величин $E_l^{(in)}$, которые характеризуют внутри-доменные взаимодействия между спинами:

$$E(t) = E^{(d)}(t) + \sum_{l=1}^n E_l^{(in)}(t), \quad \text{где } E_l^{(in)}(t) = - \sum_{i=(l-1)k+1}^{lk} \sum_{j=(l-1)k+1}^{lk} J_{ij} s_i(t) s_j(t). \quad (6)$$

Доменная динамика определяется следующим образом [3],[4]: в случайном порядке перебирают n доменов, и если для l -го домена выполняется неравенство

$$F_l(t) = \sum_{i=(l-1)k+1}^{lk} s_i(t)h_i^{(d)}(t) \geq 0,$$

то домен оставляют неизменным; если же $F_l(t) < 0$, то в следующий момент времени l -й домен переворачивают – всем входящим в него спином одновременно присваивают противоположные значения:

$$s_i(t+1) = \begin{cases} s_i(t), & \text{если } F_l(t) \geq 0, \\ -s_i(t), & \text{если } F_l(t) < 0, \end{cases} \quad \forall i \in [1+(l-1)k, lk].$$

Сравнивая последнее выражение с аналогичным выражением (3) для случайной динамики, убеждаемся в том, что доменная динамика превращается в стандартную случайную динамику, когда длина доменов равна 1: $k=1$.

Из выражения (5) следует, что доменная динамика сопровождается монотонным понижением доменной энергии состояния: если в момент времени t переворачивают l -й домен, то доменная энергия понижается на $4|F_l(t)|$:

$$E^{(d)}(t+1) = E^{(d)}(t) - 4|F_l(t)|.$$

Заметим – и это очень важно, – что одновременно ровно на такую же величину понижается энергия $E(\mathbf{s}(t))$. Это следует из выражения (6) и того очевидного факта, что одновременный переворот всех спинов l -го домена не меняет значения l -й внутри-доменной энергии $E_l^{(in)}$.

Поскольку доменная энергия состояния не может понижаться бесконечно, рано или поздно система окажется в *доменном* локальном минимуме. В этом состоянии для всех доменов будет выполняться $F_i(t) \geq 0$, и эволюция доменной динамики прекратится. Однако доменный локальный минимум совсем не обязательно будет минимумом функционала (1). Поэтому домены здесь «размораживаются» и переходят к стандартной случайной динамике. Система получает возможность опуститься по энергии еще глубже, в минимум функционала (1).

Основные надежды, которые возлагают на последовательное использование сначала доменной динамики, а затем – случайной, связаны со следующими соображениями. Энергетическую поверхность функционала (1) можно представить как своего рода овражистый ландшафт в многомерном пространстве, склоны которого изрыты огромным числом мелких локальных минимумов. Минимизация (1) осуществляется в результате скатывания по этому склону динамической системы, запущенной из некоторой начальной точки. При использовании случайной динамики два последовательных состояния системы отличаются друг от друга противоположным значением всего одной бинарной координаты – той, которую переворачивают на данном такте эволюции. Понятно, что система при этом имеет много шансов застрять в первом попавшемся локальном минимуме, даже и неглубоком. Напротив, при использовании доменной динамики два последовательных состояния системы отличаются друг от друга противоположными значениями сразу нескольких спиновых координат. Доменную динамику следует уподобить скатыванию вниз по склону более «крупным шагом». Можно ожидать, что такой более крупный шаг позволит системе миновать множество мелких локальных минимумов, в которых она застряла бы при случайной динамике.

Для простоты все определения давались выше в предположении, что домены имеют одинаковую длину k , а объединяемые в домен спины имеют последовательные номера. Понятно, что в общем случае домены могут иметь разные длины k_i , и объединять в домен можно любые спины. В связи с этим возникают вопросы: влияет ли тот или иной состав доменов на результаты минимизации? Каким образом следует формировать домены? Должны ли они быть постоянными для данной матрицы \mathbf{J} , или их следует формировать случайно на каждом такте эволюции? Можно ли здесь опереться на какие-либо общие соображения? В следующих разделах мы постараемся ответить на эти вопросы.

Замечание. По-видимому, идея доменной динамики не раз выдвигалась различными авторами. При подготовке настоящей публикации мы установили, что доменная динамика эквивалентна введенной в [6] *block-sequential dynamics*. Отметим, что *block-sequential dynamics* изучалась в [6] применительно к задаче увеличения емкости памяти модели Хопфилда. Насколько нам известно, использование доменной динамики для минимизации функционала (1) впервые изучалось именно в работах [3]-[5].

3. Кластерный принцип формирования доменов

1. Хеббовская матрица специального вида. Для модели Хопфилда известна ситуация, когда домены естественно порождаются особенностями устройства матрицы связей \mathbf{J} [7]. Приведем соответствующие результаты, расставляя нужные нам акценты. В конце раздела будет сформулирован общий принцип формирования доменов.

Пусть имеется N M -мерных векторов-столбцов \mathbf{x}_i с бинарными координатами: $x_i^{(\mu)} = \pm 1$, $i = 1, \dots, N$, $\mu = 1, \dots, M$. Векторы \mathbf{x}_i нумеруются нижним индексом $i \in [1, N]$, а их координаты – верхним индексом $\mu \in [1, M]$; соотношение между размерностью векторов (M) и их числом (N) для наших целей значения не имеет.

Образуем $(M \times N)$ -матрицу \mathbf{X} , столбцами которой являются M -мерные векторы \mathbf{x}_i :

$$\mathbf{X} = \begin{pmatrix} x_1^{(1)} & x_2^{(1)} & \dots & x_N^{(1)} \\ x_1^{(2)} & x_2^{(2)} & \dots & x_N^{(2)} \\ \dots & \dots & \dots & \dots \\ x_1^{(M)} & x_2^{(M)} & \dots & x_N^{(M)} \end{pmatrix}.$$

В теории нейронных сетей \mathbf{X} называют *матрицей паттернов* [2]. С ее помощью строится Хеббовская матрица связей – $(N \times N)$ -матрица скалярных произведений векторов-столбцов \mathbf{x}_i :

$$J_{ij} = \frac{(1 - \delta_{ij})}{M} (\mathbf{x}_i, \mathbf{x}_j) = \frac{(1 - \delta_{ij})}{M} \sum_{\mu=1}^M x_i^{(\mu)} x_j^{(\mu)}, \quad 1 \leq i, j \leq N.$$

Поскольку длины векторов-столбцов \mathbf{x}_i равны \sqrt{M} , матричные элементы J_{ij} суть косинусы углов между соответствующими векторами: $|J_{ij}| \leq 1$; на главной диагонали матрицы стоят нули.

Мы рассмотрим тот специальный случай, когда число n различных векторов-столбцов у матрицы \mathbf{X} меньше, чем N : $\mathbf{x}_1 \dots \neq \dots \mathbf{x}_l \dots \neq \dots \mathbf{x}_n$, $n < N$. Иными словами, каждый вектор-столбец \mathbf{x}_i матрицы \mathbf{X} может по несколько раз повторяться. Если k_i – число повторений вектора-столбца \mathbf{x}_i , то $\sum_1^n k_i = N$. Без ограничения общности можно считать, что первые k_1 векторов-столбцов матрицы \mathbf{X} совпадают между собой, следующие k_2 векторов-столбцов совпадают между собой, и так далее – последние k_n векторов-столбцов тоже совпадают:

$$\mathbf{X} = \left(\begin{array}{ccc|ccc|ccc} x_1^{(1)} & \dots & x_1^{(1)} & x_2^{(1)} & \dots & x_2^{(1)} & \dots & x_n^{(1)} & \dots & x_n^{(1)} \\ x_1^{(2)} & \dots & x_1^{(2)} & x_2^{(2)} & \dots & x_2^{(2)} & \dots & x_n^{(2)} & \dots & x_n^{(2)} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ x_1^{(M)} & \dots & x_1^{(M)} & x_2^{(M)} & \dots & x_2^{(M)} & \dots & x_n^{(M)} & \dots & x_n^{(M)} \end{array} \right), \quad k_l \geq 1. \quad (7)$$

Ясно, что соответствующая Хеббовская матрица будет состоять из $n \times n$ блоков. Размерность (lm) -го блока равна $(k_l \times k_m)$, а все его элементы равны одному и тому же числу J_{lm} – косинусу угла между векторами \mathbf{x}_l и \mathbf{x}_m ($l, m = 1, \dots, n$). На главной диагонали матрицы \mathbf{J} стоят квадратные $(k_l \times k_l)$ -блоки, состоящие из сплошных 1 (за исключением диагональных матричных элементов, которые по определению равны 0):

$$\mathbf{J} = \left(\begin{array}{ccc|ccc|ccc} \overbrace{0 \ 1 \ 1}^{k_1} & & & \overbrace{J_{12} \ \dots \ J_{12}}^{k_2} & & & \overbrace{J_{1n} \ \dots \ J_{1n}}^{k_n} & & & \\ 1 & \ddots & 1 & \vdots & J_{12} & \vdots & \dots & \vdots & J_{1n} & \vdots \\ 1 & 1 & 0 & J_{12} & \dots & J_{12} & & J_{1n} & \dots & J_{1n} \\ J_{21} & \dots & J_{21} & 0 & 1 & 1 & & J_{2n} & \dots & J_{2n} \\ \vdots & J_{21} & \vdots & 1 & \ddots & 1 & \dots & \vdots & J_{2n} & \vdots \\ J_{21} & \dots & J_{21} & 1 & 1 & 0 & & J_{2n} & \dots & J_{2n} \\ \vdots & & \vdots & & & \ddots & & \vdots & & \vdots \\ J_{n1} & \dots & J_{n1} & J_{n2} & \dots & J_{n2} & & 0 & 1 & 1 \\ \vdots & J_{n1} & \vdots & \vdots & J_{n2} & \vdots & \dots & 1 & \ddots & 1 \\ J_{n1} & \dots & J_{n1} & J_{n2} & \dots & J_{n2} & & 1 & 1 & 0 \end{array} \right), \quad |J_{lm}| < 1, \quad l, m \in [1, n]. \quad (8)$$

Оказывается [7], что подобное устройство матрицы связей накладывает достаточно жесткие ограничения на вид конфигураций, доставляющих локальные минимумы функционалу (1). А именно, локальный минимум непременно должен иметь «кусочно-постоянный» вид:

$$\mathbf{s} = (\underbrace{s_1 \ \dots \ s_1}_{k_1}, \underbrace{s_2 \ \dots \ s_2}_{k_2}, \dots, \underbrace{s_n \ \dots \ s_n}_{k_n}), \quad s_l = \pm 1, \quad l = 1, \dots, n. \quad (9)$$

Иными словами, первые k_1 координат локального минимума должны быть одинаковыми, следующие k_2 координат должны быть равны друг другу, и так далее – последние k_n координат должны быть равны друг другу. Блоки «знакопостоянства» в выражении (9) определяются блочной структурой Хеббовской матрицы (8); или (что есть то же самое) – структурой одинаковых столбцов матрицы паттернов \mathbf{X} (7). Доказательство того, что локальные минимумы (1) непременно имеют вид (9), основано на том, что спины из одного блока имеют одинаковые

связи с остальными спинами, и под действием локального поля ведут себя одинаково (доказательство приведено в п.1 Приложения).

Не все конфигурации вида (9) будут локальными минимумами функционала (1), но искать минимумы надо среди таких конфигураций. Иначе говоря, с точки зрения минимизации $E(s)$ бессмысленно рассматривать конфигурации с координатами разных знаков *внутри* блоков знакопостоянства. Имеет смысл рассматривать только конфигурации вида (9). Все спины из одного блока знакопостоянства либо одновременно удовлетворены, либо нет. В последнем случае не имеет никакого смысла переворачивать неудовлетворенные спины по отдельности – ведь при этом возникнут конфигурации с координатами разных знаков внутри блока знакопостоянства, а рассматривать такие конфигурации просто бессмысленно. И оказывается, что при матрице связей вида (8) можно одновременно переворачивать целый блок неудовлетворенных спинов – энергия (1) будет при этом понижаться. Иными словами, блоки знакопостоянства в (9) выступают здесь как *естественные* домены.

2. Кластерный принцип формирования доменов. Вообще говоря, и для блочно-постоянной матрицы связей (8) можно сформировать домены случайным образом, а затем использовать их для доменной минимизации функционала (1). Экспериментальная часть настоящей работы как раз и посвящена исследованию того, какой из способов формирования доменов лучше. Но прежде чем переходить к описанию экспериментов сформулируем в двух пунктах общий принцип формирования «правильных доменов»:

а) в один домен объединяют спины, связанные друг с другом сильнее, чем с другими спинами;

б) значения спинов, объединенных в один домен, устанавливаются одинаковыми.

Этот рецепт естественно назвать *кластерным принципом* формирования доменов из-за его очевидной связи с задачей кластеризации симметричной матрицы [8]. В этой задаче требуется привести матрицу к блочно-диагональному виду таким образом, чтобы элементы внутри диагональных блоков превосходили матричные элементы вне диагональных блоков. Более подробно мы обсудим эту задачу в конце статьи.

Мотивировка кластерного принципа вполне очевидна. Матричный элемент J_{ij} является мерой связи между i -м и j -м спинами. Кластерный принцип формирования доменов исходит из того, что два сильно связанных спина будут, скорее всего, сходным образом взаимодействовать и с остальными спинами. Тогда локальные поля, действующие на i -й и j -й спины, будут иметь близкие значения. Естественно предположить, что под действием одинаковых локальных полей спины будут вести себя одинаково. Вопрос о том, какие спины можно считать «сильно связанными», обсуждается в следующем разделе.

В заключение заметим, что Хеббовская матрица (8) отвечает предельно идеализированной ситуации – ее диагональные блоки состоят из 1. Формируя «правильные домены» согласно (9), мы будем объединять в один домен спины, связанные друг с другом предельно сильно. Такие матрицы редко встречается на практике. В компьютерных экспериментах мы исследовали не только идеализи-

рованный случай предельно сильных внутри-доменных связей, но и более реалистичные матрицы.

4. Компьютерное моделирование

1. Сильно связанные группы спинов. Первая серия экспериментов проводилась для Хеббовских матриц (8) с предельно сильными внутри-групповыми связями – см. п.1 предыдущего раздела. Внешние параметры задачи были равны следующим значениям: число векторов-столбцов матрицы паттернов $N=1000$, их размерность $M=60$, число доменов $n=40$, размеры доменов k_i – случайные числа из интервала $[1, 45]$: $\sum_{i=1}^{40} k_i = 1000$. Координаты $x_i^{(\mu)}$ в матрице паттернов (7) принимали значения ± 1 с равными вероятностями. Как результат, на главной диагонали Хеббовской матрицы располагались 40 $(k_i \times k_i)$ -блоков, состоящих из одних единиц: $J_{ii}^{(in)} = 1$. Элементы вне-диагональных блоков суть случайные величины со средним значением 0 и дисперсией $1/M$: $\langle J_{lm}^{(out)} \rangle = 0$, $\sigma(J_{lm}^{(out)}) = 1/\sqrt{M}$. Кластерный принцип дает здесь 40 доменов кусочно-постоянного вида (9), образованных сорока группами сильно связанных спинов.

Всего генерировалось 200 подобных матриц, и для каждой матрицы функционал (1) минимизировали с использованием 3 динамических сценариев:

а) RANDOM: случайную динамику запускали с 1000 случайных стартовых конфигураций;

б) DM-RND (случайные домены): с тех же самых случайных конфигураций запускали доменную динамику, сформировав $n=40$ случайных доменов одинаковой длины $k=25$; в домены объединяли по 25 спинов со случайными номерами, а спинам внутри домена случайным образом присваивали значения ± 1 ;

в) DM-CLS (кластерные домены): с 1000 случайных стартовых конфигураций блочно-постоянного вида (9) запускали доменную динамику.

По достижении доменного минимума домены «размораживали», и система (уже в случайной динамике) опускалась в более глубокий минимум функционала (1)². Таким образом, для каждой матрицы получали 3000 локальных минимумов функционала (1). Затем определяли самый глубокий из них, и вычисляли частоту его отыскания в каждом из трех динамических сценариев: RANDOM, DM-RND и DM-CLS. Тот тип динамики, для которого частота отыскания самого глубокого минимума будет наибольшей, следует, очевидно, считать наилучшим.

На рис.1 для всех трех динамических сценариев приведены результаты, усредненные по 200 случайным испытаниям. По оси абсцисс даны 3 типа дина-

² В случае кластерных доменов вида (9) доменный локальный минимум является и локальным минимумом функционала (1) – см. п.2 Приложения; в этом случае размораживание доменов не дает дополнительного выигрыша по энергии.

мики, по оси ординат – средняя частота отыскания самого глубокого минимума (выраженная в процентах). Мы видим, что доменная динамика с кластерными доменами (DM-CLS) приводит в самый глубокий минимум в среднем **в 30 раз чаще**, чем случайная доменная динамика (DM-RND) или случайная динамика (RANDOM). Преимущества кластерного принципа формирования доменов здесь очевидны.

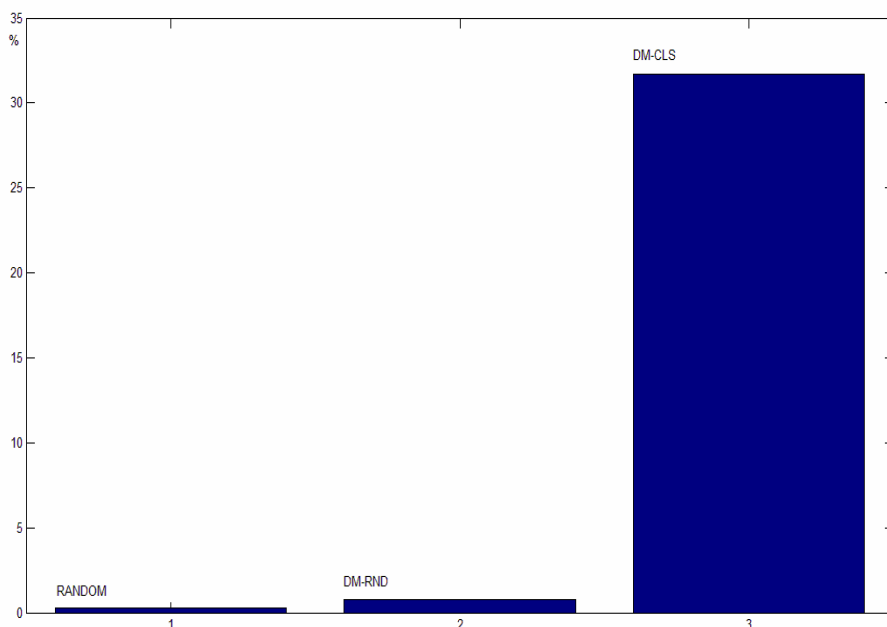


Рис.1. Средняя частота отыскания самого глубокого локального минимума для всех трех динамик.

2. Ослабление связей внутри сильно связанных спиновых групп. В описанных экспериментах связи между спинами внутри «правильных» групп были равны максимально возможному значению $J_{ll}^{(in)} = 1$, поскольку у матрицы паттернов (7) все k_l векторов-столбцов l -й группы были одинаковыми. Что произойдет, если эти группы векторов слегка «размыть»? Если векторы-столбцы внутри «правильных» групп будут не одинаковыми, а слегка отличающимися друг от друга? Как это отразится на результатах минимизации?

Описанные эксперименты были повторены с «размытыми» группами векторов-столбцов x_l . Теперь первые k_l векторов-столбцов матрицы (7) брались не идентичными друг другу, а возникали в результате случайного мультипликативного искажения M -мерного вектора x_l , а именно: координаты вектора

x_1 с вероятностью b независимым и случайным образом умножались на «-1». Следующие k_2 векторов-столбцов матрицы X получались в результате аналогичного искажения M -мерного случайного вектора x_2 , и так далее – последние k_n векторов-столбцов матрицы X являлись искажениями M -мерного случайного вектора x_n .

Теперь связи между спинами внутри «правильных» групп уже не равны тождественно 1, а являются случайными величинами со средним значением $\langle J_{ij}^{(in)} \rangle = (1 - 2b)^2$. Параметр искажения b характеризует степень неоднородности «правильных» групп спинов: чем больше b , тем меньше среднее значение внутри-групповой связи, тем менее однородными являются «правильные» группы спинов. Связи между спинами из двух различных групп по-прежнему оцениваются как $J_{ij}^{(out)} \sim \pm 1/\sqrt{M}$.

Величины b и отвечающие им средние значения внутри-групповой связи приведены в табл.1. Только для $b=0.02$ и $b=0.05$ «правильные» группы спинов можно считать сильно связанными. Действительно, среднее значение внутри-групповой связи в этих случаях равно $\langle J_{ij}^{(in)} \rangle \approx 0.9$ и $\langle J_{ij}^{(in)} \rangle \approx 0.8$ соответственно. Иными словами, углы между M -мерными векторами x_i из «правильной» группы меньше 45° ; такие группы векторов еще можно считать плотными, а соответствующие группы спинов – сильно связанными. Но уже для $b=0.1$ имеем $\langle J_{ij}^{(in)} \rangle = 0.64$, а для $b=0.2$ среднее значение внутри-групповой совсем невелико: $\langle J_{ij}^{(in)} \rangle = 0.36$. При таком среднем значении матричного элемента то, что мы по инерции продолжаем считать «правильной» группой спинов, является объединением нескольких слабо связанных подгрупп. Объединяя в один домен несколько слабо связанных спиновых групп, мы, фактически, формируем случайные домены. Можно ожидать, что с ростом b результаты для DM-CLS-динамики будут все больше походить на результаты для DM-RND-динамики.

Таблица 1. Среднее значение внутри-групповой связи между спинами $\langle J_{ij}^{(in)} \rangle$ в зависимости от b .

b	0.02	0.05	0.1	0.2
$\langle J_{ij}^{(in)} \rangle$	0.92	0.81	0.64	0.36

На рис.2 для трех типов динамики показано, как от параметра b зависит средняя частота отыскания самого глубокого минимума. Мы видим, что с ростом b результаты для «правильной» доменной динамики (DM-CLS) все меньше отличаются от результатов для случайной доменной динамики (DM-RND) - в особенности, начиная с $b=0.1$. Что и следовало ожидать, поскольку с ростом b «правильные» группы спинов все меньше отличаются от случайных доменов.

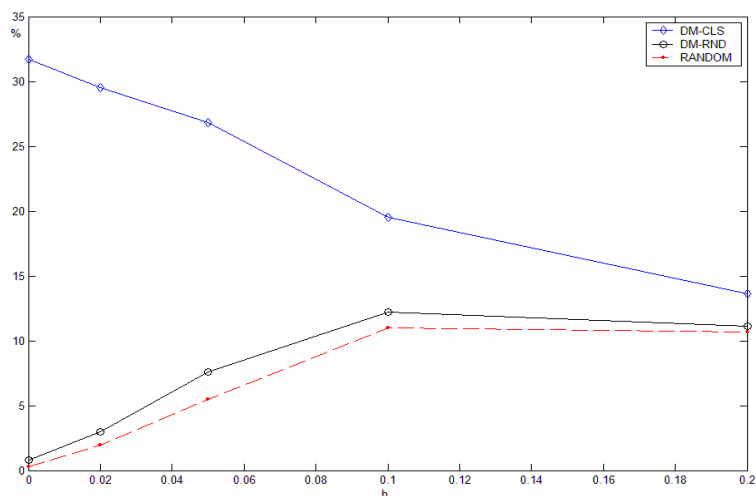


Рис.2. Средняя частота отыскания самого глубокого локального минимума для всех трех динамик как функция параметра b .

Отметим, что результаты для случайной доменной динамики лишь незначительно лучше результатов для случайной динамики: DM-RND-кривая лишь слегка превосходит RANDOM-кривую. Возможно, это объясняется тем, что длина случайных доменов в наших экспериментах ($k = 25$) далека от оптимальной длины. В работе [5] было установлено, что наилучшие результаты для доменной динамики получаются при $k=2$.

3. Вклады различных динамик в глубину локального минимума. После остановки в доменном локальном минимуме переходят к случайной динамике, и система опускается еще глубже, в локальный минимум функционала (1). Условимся под *глубиной* локального минимума понимать абсолютное значение соответствующей энергии. Пусть D – глубина доменного локального минимума, а $E \geq D$ – глубина окончательного локального минимума функционала (1), в котором система оказывается после «размораживания» доменов. Введем величины $d=D/E$ и $r=(E-D)/E$. Они неотрицательны, и их сумма равна 1:

$$0 \leq d, r \leq 1, d + r = 1.$$

Величины d и r характеризуют в относительных единицах то, какая часть глубины минимума определяется доменной динамикой, а какая – случайной динамикой. Например, если $r \approx 0$, то можно считать, что практически вся глубина локального минимума определяется доменной динамикой, а увеличение глубины за счет случайной динамики незначительно. Если же $r \approx 1$, то имеет место об-

ратная ситуация: доменная динамика почти не приводит к понижению энергии, а вся глубина локального минимума определяется случайной динамикой. Фактически, d и r характеризуют относительный вклад в глубину локального минимума доменной и случайной динамик. Интересно сопоставить друг с другом значения d - и r -характеристик. Зависит ли соотношение между ними от типа доменной динамики или нет?

Для ответа на эти вопросы мы усредняли r -характеристику сначала – по 1000 случайным стартам для каждой матрицы, а затем – по 200 случайным матрицам. Такое усреднение проделывалось для двух вариантов доменной динамики: со случайными доменами (DM-RND) и с кластерными доменами (DM-CLS). Графики усредненных r -характеристик приведены на рис.3. По оси абсцисс отложен уровень искажений b , по оси ординат – усредненная по 200000 стартов r -характеристика для кластерной доменной динамики (DM-CLS) и для случайной доменной динамики (DM-RND).

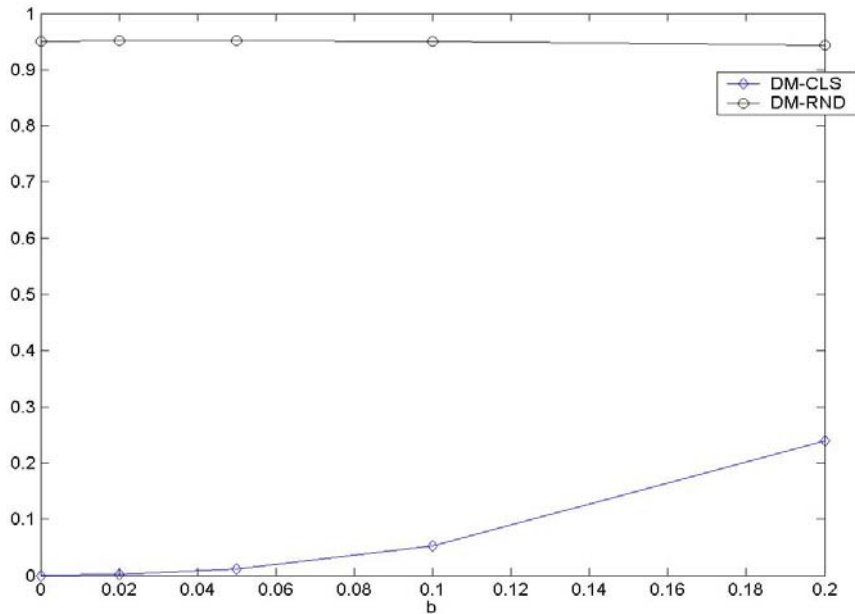


Рис. 3. Усредненные значения r -характеристики для кластерных доменов (DM-CLS) и случайных доменов (DM-RND) – см. текст.

Мы видим, что для случайной доменной динамики выполняется приближенное равенство $r \approx 1$. Иначе говоря, использование случайных доменов длины $k=25$ характеризуется тем, что в среднем 95% глубины локального минимума определяется случайной минимизацией. Доменная динамика дает весьма незна-

чительный вклад в глубину локального минимума (около 5%). Этот результат практически не зависит от уровня искажений b – см. ход (DM-RND)-кривой. Можно сказать, что использование случайных доменов длины $k=25$ лишь незначительно улучшает результаты стандартной случайной минимизации.

Напротив, при использовании кластерных доменов имеем: $r \ll 1$. Большая часть глубины локального минимума определяется здесь именно доменной динамикой, а вклад случайной динамики сравнительно мал – см. ход (DM-CLS)-кривой. При нулевом уровне искажений ($b=0$) доменные локальные минимумы вообще являются минимумами функционала (1) (см. п.2 Приложения); в этом случае выполняется точное равенство $r = 0$. По мере увеличения параметра b растет и значение r -характеристики, однако происходит это очень медленно.

5. Заключение

Мы видели, что когда в эмпирических данных можно обнаружить четко выраженные группы сильно связанных спинов, их использование в качестве доменов позволяет на порядок чаще находить глобальный минимум, чем с помощью случайных доменов. Ослабление связей внутри таких групп сильно связанных спинов, «размытие» этих групп, стирает разницу между результатами, характерными для двух способов формирования доменов. Иными словами, результаты компьютерных экспериментов свидетельствуют о продуктивности кластерного принципа формирования доменов. Число и состав доменов определяются при этом в результате кластеризации матрицы \mathbf{J} .

Под кластеризацией матрицы понимаются приведение ее к такому блочно-диагональному виду, когда матричные элементы внутри диагональных блоков превосходят матричные элементы вне этих блоков. Конкретная постановка задачи кластеризации определяется тем, в каком смысле понимать превосходство внутри-диагональных матричных элементов над вне-диагональными. Широкий диапазон различных подходов к решению каждой конкретной задачи. Обзоры методов кластеризации можно найти в [8].

Для матрицы связей корреляционного типа, когда матричные элементы являются скалярными произведениями некоторого набора векторов, хорошо зарекомендовала себя процедура кластеризации, предложенная в [9]. Эта процедура позволяет не только выделить плотные группы векторов, но и разобраться в том, какие из этих групп расположены ближе друг к другу, а какие дальше. Именно эта информация и требуется для объединения сильно связанных спинов в домены. Вычислительная сложность процедуры кластеризации $\sim O(N^2)$. Однако обобщить ее на матрицы связей общего вида – необязательно корреляционные – пока не удается.

Автор признателен А.А.Мурашкину, выполнившему для настоящей работы весь вычислительный эксперимент. Работа выполнялась в рамках программы "Фундаментальные основы информационных технологий и систем" (проект 1.8) при финансовой поддержке грантом РФФИ 06-01-00109.

Приложение

Докажем здесь два утверждения относительно локальных минимумов функционала (1) для случая, когда матрица связей имеет блочную структуру (8); эти утверждения были использованы нами выше.

1. Кусочно-постоянный вид локальных минимумов. Покажем, что первые k_1 координат локального минимума должны быть одинаковыми – см. выражение (9). Рассуждение, устанавливающее справедливость этого утверждения, проведем для первых двух координат. Пусть конфигурация $\mathbf{s} = (s_1, s_2, s_3, \dots, s_N)$ является локальным минимумом. Тогда каждая координата

s_i должна совпадать со знаком локального поля h_i : $s_1 h_1 = s_1 \sum_{j=1}^N J_{1j} s_j \geq 0$ и

$s_2 h_2 = s_2 \sum_{j=1}^N J_{2j} s_j \geq 0$. Поскольку матрица \mathbf{J} имеет вид (8), локальные поля можно

записать следующим образом: $h_1 = s_2 + H$, $h_2 = s_1 + H$. Второе слагаемое

$H = \sum_{j=3}^N J_{ij} s_j$ в обоих случаях одинаково, поскольку все матричные элементы

первой и второй строк с номерами $j > 2$ одинаковы: $J_{1j} = J_{2j}$, $j > 2$ – см. выражение (8). Итак, одновременно должны выполняться два неравенства:

$s_1 s_2 + s_1 H \geq 0$ и $s_2 s_1 + s_2 H \geq 0$. Если предположить, что координаты s_1 и s_2 различны, то $s_1 s_2 < 0$ и одновременно должно выполняться: $1 \leq H$ и $1 \leq -H$ – что невозможно. Следовательно, наше предположение неверно и координаты s_1 и s_2 совпадают. Доказательство того, что локальные минимумы функционала (1) с матрицей \mathbf{J} (8) имеют кусочно-постоянный вид (9), на этом закончено.

2. Доменные локальные минимумы кусочно-постоянного вида. Легко сообразить, что любая конфигурация вида (9), являющаяся доменным локальным минимумом, будет и локальным минимумом функционала (1). Действительно, пусть для первого домена выполняется неравенство: $F_1 = \sum_{i=1}^k s_i h_i^{(d)} \geq 0$.

Поскольку первые k_1 координат s_i (которые как раз и объединены в первый домен) равны друг другу, сумма в правой части последнего выражения есть сумма k_1 одинаковых слагаемых: $F_1 = k_1 s_1 h_1^{(d)}$; следовательно,

$s_1 h_1^{(d)} = s_1 \sum_{j=k+1}^N J_{1j} s_j \geq 0$. Используя выражение (4), связывающее локальное поле

h_i с доменным локальным полем $h_i^{(d)}$, получаем, что положительным будет и следующее выражение:

$$s_1 h_1 = s_1 h_1^{(d)} + s_1 \sum_{j=1}^{k_1} J_{1j} s_j = s_1 h_1^{(d)} + (k_1 - 1) > 0. \quad (10)$$

Это означает, что координата s_1 совпадает со знаком локального поля h_1 , действующего на нее со стороны всей сети. Иначе говоря, доменный локальный минимум будет и локальным минимумом функционала (1). Следовательно, «размораживание» здесь доменов не приведет к дальнейшему понижению энергии состояния. В терминах п.4.3 это означает, что $r=0$.

Обратное неверно: из того, что кусочно-постоянная конфигурация вида (9) является локальным минимумом функционала (1), вовсе не следует, что она будет и доменным локальным минимумом. Иначе говоря, из того, что $s_1 h_1 > 0$, совсем не следует, что $s_1 h_1^{(d)} > 0$. Это легко увидеть, обратившись к выражению (10): $s_1 h_1^{(d)}$ может быть отрицательным, но если число спинов k_1 в домене достаточно велико, сумма в правой части выражения (10) будет положительной. Иначе говоря, одновременно будут выполняться два неравенства: $s_1 h_1 > 0$, $s_1 h_1^{(d)} < 0$. Это означает, в частности, что доменная динамика позволяет иногда выбираться из не очень глубоких локальных минимумов функционала (1).

Литература

- [1] A.K. Hartmann, H. Rieger. Optimization Algorithms in Physics. Wiley-VCH, Berlin, 2001.
- [2] J. Hertz, A. Krogh, R. Palmer. Introduction to the Theory of Neural Computation. Addison-Wesley, 1991.
- [3] Б.В. Крыжановский, Б.М. Магомедов, А.Л. Микаэлян. Доменная модель нейронной сети. Доклады Академии наук. 2005. Т. 401. №4. С. 462-466.
- [4] Б.В. Крыжановский, Б.М. Магомедов. Доменный алгоритм релаксации нейронной сети и оптимизация квадратичного функционала. Труды "МСО-2005". М.: МГУ, 2005. С. 607-613.
- [5] V. Kryzhanovsky, B. Magomedov. On the probability of finding local minima in optimization problems. *Proceedings of IJCNN'2006*, Vancouver: 2006. PP. 5882–5887.
- [6] A.V.M. Herz, C.M. Marcus. Distributed dynamics in neural networks. *Phys. Rev. E*47. 1993. PP.2155-2161.
- [7] Л.Б. Литинский. О прямом вычислении неподвижных точек нейронной сети. Теоретическая и математическая физика. 1994. Т.101(3). С. 467-479.
- [8] Н.Г. Загоруйко. Прикладные методы анализа данных и знаний. Новосибирск: Изд-во Ин-та математики, 1999.
- [9] L. Litinskii, D. Romanov. Neural network clustering based on distances between objects. St. Kollias, A. Stafylopatis, W Duch, E.Oja (Eds.). *Artificial Neural Networks – ICANN 2006*. 16h International Conference, Athens, Greece, September 10-14, 2006, Proceedings, Part II. PP.437-443.

Статья поступила 06 февраля 2007 г.
После доработки 29 августа 2007 г.