

Структурный анализ поведенческой динамики*

Темлянец А. В., Ветров Д. П., Кропотов Д. А.

alexander.temlyantsev@gmail.com

Москва, ВМиК МГУ, Вычислительный Центр РАН

В данной работе построено решение задачи анализа поведения в рамках конкретного биологического эксперимента методами структурного распознавания. Предлагаемый подход формализует взаимосвязи между поведением и эмоциональной динамикой живой системы в виде *графической модели*, в рамках которой удается построить эффективные алгоритмы обучения и распознавания.

Развитие биологических представлений об организме как о живой системе явилось стимулом для возникновения целого ряда аналогий и применения методов биологической науки в смежных дисциплинах. Преимущество изучения целостного организма с особой силой выражено уже в очень ранних работах И. П. Павлова. Так, например, еще в конце прошлого столетия он выдвинул идею, что «наиболее нормальные функции организма можно изучать не у ограниченного в подвижности животного, т. е. в условиях вивисекции, а у целостного, ненаркотизированного животного» [1]. Одна лишь видеозапись активности лабораторного животного содержит в себе существенную информацию о динамике обуславливающих ее скрытых функциональных систем. Масштаб и относительная простота экспериментов, вкупе с их научной значимостью обуславливают возрастающий интерес к построению систем автоматического анализа результирующих данных.

В данной работе предложены алгоритмы распознавания эмоционального состояния животного по видеозаписи на базе композитно-иерархических скрытых марковских моделей, сочетающих в себе ряд фундаментальных биологических представлений:

- Динамика поведения животного может быть представлена последовательностью поведенческих актов. Идея *поведенческого акта* достаточно точно передается соответствующим ему глаголом: «идти», «бежать», «красться», выражающим действие вкупе с целью, и по достижении которой оно будет прекращено [1].
- Эмоциональное состояние есть набор независимых элементарных эмоций, влияющих на поведение особи композитно. Особи, находящейся в сложном эмоциональном состоянии, одновременно присущи свойства, характерные для каждой активной в данный момент элементарной эмоции.

Байесовские сети

Определение 1. Пусть $G = \langle V, \Gamma \rangle$ — ориентированный граф с множеством вершин V и множе-

ством ребер Γ . Будем говорить, что случайный вектор $(x_1 \dots x_n)$, $n = |V|$ удовлетворяет графической структуре G , если существует такое взаимнооднозначное соответствие $\{x_1, \dots, x_n\} \xleftrightarrow{f} V$, что

$$p(x_1 \dots x_n) = \prod_{i=1}^{|V|} p(x_i | f^{-1}(\pi(f(x_i)))) \quad (1)$$

где $\pi(v) = \{v_{prev} : (v_{prev}, v) \in \Gamma\}$ — множество родителей вершины $v \in V$ в графе G .

Говорят, что семейство распределений удовлетворяет G , если каждое распределение этого семейства удовлетворяет G

Язык графов оказывается удобным и гибким инструментом для отражения в вероятностной модели причинно-следственных связей, присущих описываемому объекту. Зафиксировав графическую структуру вероятностной модели, в достаточном общем случае удается предложить эффективные методы решения основных задач статистического анализа сразу для всех удовлетворяющих этой структуре семейств распределений. Однако наиболее завершенных прикладных результатов удастся достичь, если ввести дополнительные аналитические ограничения, например, зафиксировав явно параметрическое семейство распределений для каждой вершины графа (например, для применения ряда критериев классической статистики необходимо потребовать нормальности элементов выборки):

Определение 2. Байесовской сетью называется система $\mathfrak{A} = \langle G = \langle V, \Gamma \rangle, \mathcal{P} \rangle$ где G есть ориентированный граф,

$$\mathcal{P} = \{(v, p) : v \in V, p = p(v | \pi(v), \theta_v)\}, \quad \theta_v \in \Theta_v,$$

причем каждой вершине $v \in V$ соответствует единственная пара $(v, p) \in \mathcal{P}$. Говорят, что случайный вектор (v_1, \dots, v_n) , $n = |V|$ удовлетворяет байесовской сети, если при некоторых $\theta_1, \dots, \theta_n$

$$p(v_1 \dots v_n) = \prod_{i=1}^{|V|} p(v_i | \pi(v_i), \theta_i). \quad (2)$$

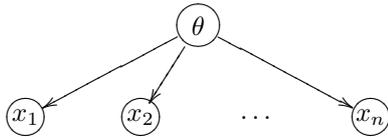
Представление (2) называется *факторизацией совместного распределения*. Отметим, что каждая

*Работа выполнена при финансовой поддержке РФФИ, проект №08-01-00405

вершина v_i входит в факторизацию, как минимум, один раз в множителе $p(v_i | \pi(v_i), \theta_i)$. При этом она может входить и в другие сомножители, являясь предком какой-либо другой вершины графа G . Будем далее говорить, что этот множитель соответствует вершине v_i .

Часто удобно заранее разбить множество вершин на наблюдаемые (observable) и скрытые (hidden), обозначая $V = \{v_1^{obs}, \dots, v_r^{obs}, v_1^{hid}, \dots, v_t^{hid}\}$.

Пример 1. В классической статистике под выборкой понимают совокупность независимых одинаково распределенных случайных величин. Такая ситуация описывается графической структурой



Здесь наблюдаемые вершины x_1, \dots, x_n соответствуют объектам выборки, а скрытая θ — неизвестному параметру распределения.

В статистическом анализе данных особую значимость имеют следующие задачи.

Задача 1. Обучение: оценить значения параметров по известным значениям наблюдаемых.

Задача 2. Восстановление значений скрытых переменных:

$$p(v_1^{obs}, \dots, v_r^{obs}, v_1^{hid}, \dots, v_t^{hid} | \theta) \rightarrow \max_{v_1^{hid}, \dots, v_t^{hid}}.$$

Следует упомянуть и задачу выбора наиболее правдоподобной структуры графа (задача выбора графической модели), однако в существующих сегодня постановках в этом направлении еще не получено значительных результатов; предложенные методы ее решения носят, как правило, характер перебора.

Построение графической модели поведенческой динамики

Коротко проследим ряд основных этапов на пути к цели исследования — построению эмоциональной динамики животного по видеозаписи его активности. Каждая вводимая здесь случайная переменная войдет как вершина в итоговую графическую модель. Таким образом, можно говорить об итоговом совместном распределении.

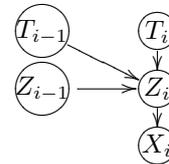
1. Устройство видеозаписи в клетке выдает последовательность кадров-изображений I_1, \dots, I_N , на основании которой строится признаковое описание активности животного X_1, \dots, X_N , представляющее собой набор механических характеристик его движения (производные координат центра масс и других характерных точек — носа, хвоста и т. п.).

2. Предполагается, что X_1, \dots, X_N представляют собой реализации случайных величин. Каждая случайная величина X_i зависит от активного в i -й момент времени поведенческого акта, номер которого в свою очередь полагается реализацией случайной величины Z_i :



Соответствующий X_i множитель итогового совместного распределения имеет вид $p(x_i | z_i) = \Phi(x_i, \mu(z_i), \Sigma(z_i))$, где $\Phi(x, \mu, \Sigma)$ есть функционал плотности нормального закона с математическим ожиданием μ и матрицей ковариации Σ .

3. Z_i зависит от активного поведенческого акта в предыдущий момент, а также от локальной эмоциональной динамики (см. следующий пункт):



Ввиду того, что множества возможных поведенческих актов и эмоциональных состояний конечны (см. следующий пункт), соответствующий z_i множитель итогового распределения представляет собой функцию дискретных переменных

$$a_{\zeta|\eta, \sigma, \tau}^i = p(z_i = \zeta | z_{i-1} = \eta, T_i = \sigma, T_{i-1} = \tau).$$

Здесь и далее $\zeta, \eta \in \{1, \dots, k\}$, σ, τ — булевы векторы размерности l , σ_i — булев вектор размерности l с единицей в i -й позиции и нулями в остальных.

4. Эмоциональное состояние в каждый момент времени моделируется бинарным вектором $T_i = (t_i^1, \dots, t_i^l)$, где $t_i^k = 1$ если и только если k -я элементарная эмоция активна в i -й момент времени. Эти переменные в совокупности и определяют эмоциональную динамику — восстановление их значений по заданным значениям наблюдаемых решит исходную задачу. Предполагается, что t_i^k зависит от t_{i-1}^k , то есть соответствующий каждому t_i^k множитель итогового распределения есть

$$b_{p,q}^{i,k} = p(t_i^k = p | t_{i-1}^k = q), \quad p, q \in \{0, 1\}.$$

5. Предполагается, что продолжительность видеосъемки достаточно мала для того, чтобы считать условные распределения введенных величин не зависящими от времени, что мотивирует *ограничения гомогенности*: для любых $i, j \in \{2, \dots, N\}$, $k \in \{1, \dots, l\}$, $p, q \in \{0, 1\}$, $\zeta, \eta, \sigma, \tau$ имеет место

$$\begin{cases} a_{\zeta|\eta, \sigma, \tau}^i = a_{\zeta|\eta, \sigma, \tau}^j = a_{\zeta|\eta, \sigma, \tau}; \\ b_{p,q}^{i,k} = b_{p,q}^{j,k} = b_{p,q}^k; \\ \mu(z_i = \zeta) = \mu(z_j = \zeta); \\ \Sigma(z_i = \zeta) = \Sigma(z_j = \zeta). \end{cases} \quad (3)$$

6. Предполагается, что выполнена *гипотеза композитности*: если в данный момент времени, не являющийся моментом смены эмоционального состояния, несколько элементарных эмоций активны одновременно, то поведению животного в одинаковой степени присущи свойства каждой из них по отдельности. Это соображение можно моделировать следующим образом: если в данный момент времени активны k элементарных эмоций, то для каждого $i = 1, \dots, k$ особь с вероятностью $1/k$ ведет себя как если бы активна была только одна i -я:

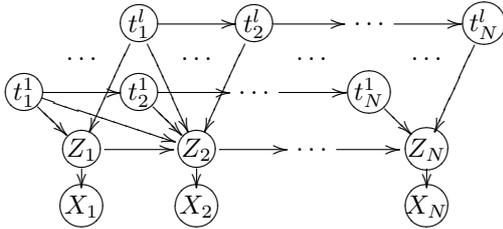
$$a_{\zeta|\eta,\sigma,\sigma} = \frac{1}{|N(\sigma)|} \sum_{i \in N(\sigma)} a_{\zeta|\eta,\sigma_i,\sigma_i}. \quad (4)$$

В момент смены эмоционального состояния поведение особи не зависит от предыстории:

$$a_{\zeta|\eta,\sigma,\tau} = a_{\zeta|\sigma}^0. \quad (5)$$

Подытожим сказанное в определении.

Определение 3. Композитно-иерархической скрытой марковской моделью (КИСММ) эмоциональной динамики будем называть байесовскую сеть с графом



дополненную ограничениями гомогенности (3) и композитности (4,5)

Отметим, что задача восстановления эмоциональной динамики по наблюдаемым автоматически решается с решением задачи 2 для КИСММ как байесовской сети, что на практике требует известных значений параметров модели a, b, μ, σ . Параметры, в свою очередь, могут быть восстановлены как решения для КИСММ задачи 1.

Задача определения значений скрытых переменных

Приведем формальную постановку задачи 2 для КИСММ:

Задача 3. $p(Z, T, X | \theta) \rightarrow \max_{Z, T}$.

Учет графической структуры модели позволяет существенно упростить процедуру максимизации функционала плотности.

Задача 3 эквивалентна следующей

$$\log p(Z, T, X) \rightarrow \max_{Z, T}. \quad (6)$$

Согласно определению модели справедливо представление:

$$\begin{aligned} \log p(Z, T, X) &= \\ &= \log p(z_1 | T_1) + \log p(T_1) + \log p(x_1 | z_1) + \\ &+ \sum_{i=2}^N \left(\log p(z_i | z_{i-1}, T_i, T_{i-1}) + \right. \\ &\quad \left. + \log p(T_i | T_{i-1}) + \log p(x_i | z_i) \right). \end{aligned}$$

Введем следующие обозначения:

$$\begin{aligned} \mathfrak{C}(z_1, T_1) &= \log p(T_1) + \log p(z_1 | T_1) + \log p(x_1 | z_1); \\ \mathfrak{C}(z_n, T_n) &= \max_{z_1, \dots, z_{n-1}, T_1, \dots, T_{n-1}} \left(\mathfrak{C}(z_1, T_1) + \right. \\ &\quad \left. + \sum_{i=2}^n \log p(z_i, T_i, x_i | z_{i-1}, T_{i-1}) \right). \end{aligned} \quad (7)$$

Опираясь только на дистрибутивность операции мах относительно сложения и умножения, можно обосновать справедливость следующего утверждения:

Утверждение 1. Для всех $n = 1, \dots, N$

$$\mathfrak{C}(z_n, T_n) = \max_{z_{n-1}, T_{n-1}} \left(\mathfrak{C}(z_{n-1}, T_{n-1}) + p(z_n, T_n, x_n | z_{n-1}, T_{n-1}) \right). \quad (8)$$

Утверждение 1 позволяет вычислять $\mathfrak{C}(z_n, T_n)$ итерационно, существенно упрощая процедуру максимизации.

Обозначим

$$\mathfrak{B}(z^*k, T_k^*) = \arg \max_{z_{n-1}, T_{n-1}} \left(\mathfrak{C}(z_{n-1}, T_{n-1}) + p(z_n, T_n, x_n | z_{n-1}, T_{n-1}) \right). \quad (9)$$

Теорема 2.

$$\max_{Z, T} p(Z, T, X) = \max_{z_N, T_N} \mathfrak{C}(z_N, T_N).$$

Пусть z^*, T^* таковы, что

$$(z_N^*, T_N^*) = \arg \max_{z_N, T_N} \mathfrak{C}(z_N, T_N); \quad (10)$$

$$(z_k^*, T_k^*) = \mathfrak{B}(z^*k + 1, T_{k+1}^*); \quad (11)$$

тогда

$$\max_{Z, T} \log p(Z, T, X) = p(z^*, T^*, X).$$

Последние два утверждения мотивируют эффективный алгоритм нахождения максимума в задаче 4. Алгоритм 1 представляет собой обобщение

Алгоритм 1. Восстановление значений скрытых переменных.

- 1: вычислить $\mathfrak{C}(z_1, T_1)$ по формуле (7);
- 2: для $i := 2, \dots, N$
- 3: вычислить $\mathfrak{C}(z_k, T_k)$, $\mathfrak{B}(z_k, T_k)$ по формулам (8), (9); все требуемые для вычисления величины уже определены на предыдущих шагах;
- 4: $(z_N^*, T_N^*) := \arg \max_{z_N, T_N} \mathfrak{C}(z_N, T_N)$;
- 5: для $i := 1, \dots, N - 1$
- 6: $(z_{N-i}^*, T_{N-i}^*) := \mathfrak{B}(z_{N-i+1}^*, T_{N-i+1}^*)$;

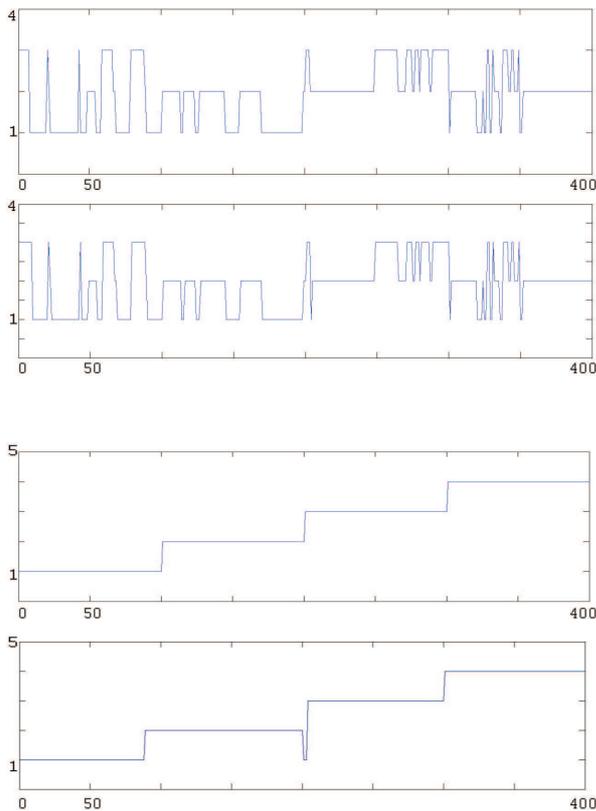


Рис. 1. Результат модельного тестирования Алгоритма 1. Сверху вниз: эмоциональная динамика, сгенерированные значения; эмоциональная динамика, ответ алгоритма; динамика поведенческого акта, сгенерированные значения; динамика поведенческого акта, ответ алгоритма.

известного алгоритма Витерби, решающего ту же самую задачу для скрытых марковских моделей.

Данный алгоритм был протестирован на модельных данных: по фиксированному θ генерировались значения скрытых и наблюдаемых переменных T_0, Z_0, X_0 . Значения θ и наблюдаемых подавались на вход алгоритма. Результат работы последнего сравнивался с Z_0, T_0 (см. рис. 1.)

Задача обучения без учителя

Построение эффективного метода настройки параметров КИСММ, подаваемых далее вместе со

значениями наблюдаемых на вход алгоритма 1, позволяет создать автономную программную систему поддержки биологического эксперимента. Естественно возникают два подхода к решению этой проблемы:

1. Получить соответствующие некоторому вектору наблюдаемых значения скрытых переменных из экспертной разметки и решить задачу $p(X, Z, T | \theta) \rightarrow \max_{\theta}$. Зафиксировав решение θ^* , применять Алгоритм 1 для произвольного вектора наблюдаемых. Теоретически, эксперту требуется один раз разметить видеозапись, чтобы далее Алгоритм 1 осуществлял разметку автоматически. Однако эксперт далеко не всегда способен провести такую разметку с достаточной для корректной работы алгоритма точностью.

2. Решать напрямую задачу обучения без учителя $p(X | \theta) \rightarrow \max_{\theta}$, что потребует явного вычисления $p(X | \theta) = \sum_{Z, T} p(X, Z, T | \theta)$ при неизвестном значении параметра θ , не допускающего использования преимуществ структурного метода.

Компромиссный подход состоит в построении итерационного процесса, способного по заданному θ_i построить новое приближение θ_{i+1} так, чтобы $p(X | \theta_{i+1}) > p(X | \theta_i)$. Выбирая в качестве начального значения θ^0 , построенное по экспертной разметке или из иных эвристических соображений, можно далее автоматически найти более точное в смысле правдоподобия значение параметра.

Означенным свойством обладает широко используемый в приложениях EM-алгоритм [2]. Соответствующая итерационная последовательность $\{\theta_i\}_{i \in \mathbb{N} \cup \{0\}}$ строится по правилу

$$Q(\theta_i, \theta_{i-1}) \rightarrow \max_{\theta_i}, i \in \mathbb{N}$$

где $Q(\theta_i, \theta_{i-1}) = \sum_Z \sum_T \ln p(X, Z, T | \theta_i) p(Z, T | \theta_{i-1})$.

Учет структуры КИСММ позволяет существенно упростить вид функционала Q и, как следствие, реализацию полученного итерационного процесса.

Литература

- [1] Анохин П. К. Очерки по физиологии функциональных систем // Москва, «Медицина», 1975. — Стр. 17-62
- [2] Bishop C. M. Pattern Recognition and Machine Learning // Springer. 2006.
- [3] Elliott R. J., Aggoun L., Moore J. B. Hidden Markov Models: Estimation and Control // Berlin: Springer-Verlag, 1995.
- [4] Ghahramani Z., Jordan M. I. Factorial hidden Markov models // Machine Learning, 1997. — Vol. 29 — Pp. 245–275.
- [5] Fine S., Singer Y., Tishby N. The Hierarchical Hidden Markov Model: Analysis and Applications // Machine Learning, 1998. — Vol. 32 — Pp. 41–62.