ФОРМАЛЬНАЯ МОДЕЛЬ НЕЙРОНА, ОБЕСПЕЧИВАЮЩАЯ НЕПРОТИВОРЕЧИВОСТЬ ПРЕДСКАЗАНИЙ♣

Витяев Е.Е.[1,2]
[1)] Институт математики им. С.Л.Соболева СО РАН, vityaev@math.nsc.ru
[2)] Новосибирский государственный университет

We formulate the formal model of neuron that provides consistent predictions. For the consistency of predictions we use maximal specific rules, that discovered by neuron.

Введение

Нами ранее была предложена формальная модель нейрона, основанная на семантическом вероятностном выводе [1-2]. Эта модель была использована для объяснения функциональных систем [1-2] и экспериментально апробирована путем создания аниматов [3].

В данной работе показывается, что эта модель позволяет решать проблему статистической двусмысленности и предсказывать без противоречий.

Проблема статистической двусмысленности состоит в том, что в процессе обучения (индуктивного вывода) мы можем получать вероятностные правила, из которых выводится противоречие. Эта проблема возникает для большинства методов машинного обучения. Пример: наблюдая людей, можно вывести два правила: если человек философ, то он не миллионер, а если он держатель приисков, то миллионер. Применяя эти два правила к известному философу П. Суппесу, мы получим, что, поскольку он философ, то он не миллионер, а, поскольку он держатель приисков, то миллионер. Получим противоречие.

Чтобы избавиться от противоречий Гемпель [4] ввел требование максимальной специфичности. В нашем примере максимально специфичными должны быть правила: если человек философ, но не держатель приисков, то он с ещё большей вероятностью не миллионер, а, если он держатель приисков, но не философ, то он также с ещё большей вероятностью миллионер. Применение этих двух правил уже не приводит к противоречиям. Максимально специфические правила должны использовать всю доступную информацию.

Полученные результаты

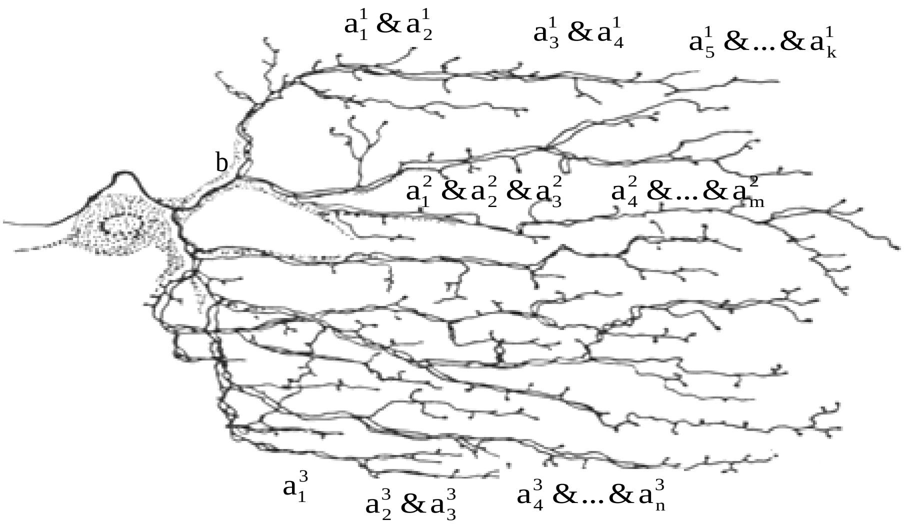
Приведём описание формальной модели нейрона на описательном уровне, ссылаясь на точные определения, приведённые в следующем разделе.

Под *информацией* поступающей на «вход» мозга будем понимать всю воспринимаемую мозгом афферентацию: мотивационную, обстановочную, пусковую, обратную, санкционирующую, афферентацию о произведенных действиях, поступающую по коллатералям на «вход» и т. д. Из экологической теории восприятия Дж. Гибсона [5] следует, что под информацией можно понимать любую характеристику энергетического потока света, звука и т.д., поступающую на «вход» мозга.

Определим информацию, передаваемую по некоторому нервному волокну на синапсы нейрона, одноместными предикатами $P_j^i(\mathbb{a}) = (x_i(\mathbb{a}) = x_{ij})$, $j = 1,...,n_i$, где $x_i(\mathbb{a})$ – некоторая информация, x_{ij} – значение истинности в ситуации (на объекте) $\mathbb{a}$. Если информация предается на возбуждающий синапс, то она воспринимается нейроном как информация об истинности предиката $P_j^i(\mathbb{a})$, если на тормозной синапс, то как отрицание предиката $\neg P_j^i(\mathbb{a})$.

Возбуждение нейрона в ситуации (на объекте) $\mathbb{a}$ и передачу этого возбуждения на аксон нейрона также определим одно-

♣ Работа поддержана грантом РФФИ № 11-07-00560-а; интеграционными проектами СО РАН № 3, 87, 136, а также работа выполнена при финансовой поддержке Совета по грантам Президента РФ для государственной поддержки ведущих научных школ (проект НШ-276.2012.1)



местным предикатом $P_0(\mathbb{a})$. Если нейрон тормозится в ситуации $\mathbb{a}$, то определим эту ситуацию как прогнозирование отрицания предиката $\neg P_0(\mathbb{a})$.

Известно, что каждый нейрон имеет рецептивное поле, возбуждающее его безусловно. Первоначальной (до всякого обучения) семантикой предиката P_0 является данное рецептивное поле. В процессе обучения эта информация обогащается и может дать достаточно специализированный нейрон типа «нейрон Билла Клинтона».

Мы предполагаем, что формирование условных правил (связей) на уровне нейрона происходит по правилу Хебба [5]. Нейрофизиологические подтверждения этого правила можно найти в работе [7].

В отличии от других формализаций, мы формализуем правило Хебба семантическим вероятностным выводом, приводимым далее.

Предикаты $P_j^i(\mathbb{a})$, $P_0(\mathbb{a})$ и их отрицания $\neg P_j^i(\mathbb{a})$, $\neg P_0(\mathbb{a})$ являются литералами, которые обозначим как $a, b, c, \ldots \in L$.

В процессе семантического вероятностного вывода нейрон обнаруживает множество правил $\{R\}$ (условных связей) вида:

$$R = (a_1 \& \ldots \& a_k \Rightarrow b),\ a_1, \ldots, a_k, b \in L, \qquad (1)$$

где $a_1, \ldots, a_k$ – возбуждающиеся (тормозные) предикаты (стимулы), приходящие на вход нейрона, а b – предикат $P_0(\mathbb{a})$ или $\neg P_0(\mathbb{a})$ аксона нейрона.

Определим способ вычисления условной вероятности правил $(a_1 \& \ldots \& a_k \Rightarrow b)$. Подсчитаем число случаев $n(a_1, \ldots, a_k, b)$, когда произошло событие $\langle a_1, \ldots, a_k, b \rangle$ – одновременное возбуждение/торможение входов $\langle a_1, \ldots, a_k \rangle$ нейрона и самого нейрона непосредственно перед действием подкрепления (которое может быть, как положительным, так и отрицательным и осуществляться, как санкционирующей афферентацией так и эмоциями [1,3]).

Среди случаев $n(a_1, \ldots, a_k, b)$ подсчитаем случаи $n^+(a_1, \ldots, a_k, b)$, когда подкрепление было положительным, а также число случаев $n^-(a_1, \ldots, a_k, b)$, когда подкрепление было отрицательным. Условную (эмпирическую) вероятность правила $(a_1 \& \ldots \& a_k \Rightarrow b)$ определим следующим образом:

$$\mu(b / a_1, \ldots, a_k) = \frac{n^+(a_1, \ldots, a_k, b) - n^-(a_1, \ldots, a_k, b)}{n(a_1, \ldots, a_k, b)}.$$

Если эта вероятность становится отрицательной, то это означает торможение нейрона с вероятностью, взятой с обратным знаком.

Формализация процесса замыкания условных связей на уровне нейрона (правило Хебба) осуществляется семантическим вероятностным выводом (определение 7), который:

1) при обнаружении стимулов, позволяющих предсказывать с некоторой вероятностью возбуждение нейрона, образует условную связь в виде правила (1).

2) при обнаружении новых стимулов, позволяющих предсказывать возбуждение нейрона с ещё большей вероятностью, по сравнению с имеющейся условной связью, присоединяет их к данной условной связи. Происходит дифференциация условной связи. Формально это определяется вероятностным выводом (определение 6).

3) в правила нейрона включаются только стимулы, которые являются сигнальными, т.е. каждый стимул должен увеличи-

вать вероятность предсказания возбуждения нейрона. Формально это определяется понятием вероятностного закона (определение 4).

4) возбуждение или торможение нейрона по совокупности правил {R} осуществляется по максимально вероятным правилам. Это подтверждается тем, что в процессе выработки условных связей, а также при замыкании условных связей на уровне нейрона, скорость ответа нейрона на условный сигнал, тем выше, чем выше вероятность условной связи.

5) максимально вероятные правила одновременно являются максимально специфическими (определение 7), которые максимально учитывают имеющуюся информацию.

6) предсказание по максимально специфическим правилам, осуществляемое нейроном, в пределе непротиворечиво (см. теорему). Поэтому в процессе дифференциации условных связей нейрон обучается предсказывать без противоречий – срабатывают либо его возбуждающие максимально специфические правила, либо тормозные, но не одновременно.

7) на рис. 1 схематически показано несколько семантических вероятностных выводов, осуществляемых нейроном. Например, условная связь $(b \Leftarrow a_1^1 \,\&\, a_2^1)$ усиливается новыми стимулами $a_3^1 \,\&\, a_4^1$ до связи $(b \Leftarrow a_1^1 \,\&\, a_2^1 \,\&\, a_3^1 \,\&\, a_4^1)$, если стимулы $a_3^1 \,\&\, a_4^1$ увеличивают условную вероятность предсказания возбуждения нейрона b.

1. $(b \Leftarrow a_1^1 \,\&\, a_2^1) \sqsubset (b \Leftarrow a_1^1 \,\&\, a_2^1 \,\&\, a_3^1 \,\&\, a_4^1) \sqsubset$
   $\sqsubset (b \Leftarrow a_1^1 \,\&\, a_2^1 \,\&\, a_3^1 \,\&\, a_4^1 \,\&\, a_5^1 \,\&\, ... \,\&\, a_k^1)$;
2. $(b \Leftarrow a_1^2 \,\&\, a_2^2 \,\&\, a_3^2) \sqsubset$
   $(b \Leftarrow a_1^2 \,\&\, a_2^2 \,\&\, a_3^2 \,\&\, a_4^2 \,\&\, ... \,\&\, a_m^2)$;
3. $(b \Leftarrow a_1^3) \sqsubset (b \Leftarrow a_1^3 \,\&\, a_2^3 \,\&\, a_3^3)$
   $\sqsubset (b \Leftarrow a_1^3 \,\&\, a_2^3 \,\&\, a_3^3 \,\&\, a_4^3 \,\&\, ... \,\&\, a_n^3)$.

Совокупность семантических вероятностных выводов, которые обнаруживает нейрон в процессе обучения, составляет его вероятностную закономерную модель (определение 5), предсказывающую возбуждение/торможение $P_0(\text{a})$ нейрона.

Методы

Приведем формальное описание модели.

Под данными обучения *Data* будем понимать все случаи возбуждения или торможения нейрона, когда было подкрепление. Множество всех правил вида (1) обозначим через Pr.

Правило $R_1 = (a_1^1 \,\&\, a_2^1 \,\&\, ... \,\&\, a_{k_1}^1 \Rightarrow c)$ будем называть *более общим*, чем правило $R_2 = (b_1^2 \,\&\, b_2^2 \,\&\, ... \,\&\, b_{k_2}^2 \Rightarrow c)$, обозначим это как $R_1 \succ R_2$, тогда и только тогда, когда $\{a_1^1, a_2^1, ..., a_{k_1}^1\} \subset \{b_1^2, b_2^2, ..., b_{k_2}^2\}$, $k_1 < k_2$ и *не менее общим* $R_1 \gtrsim R_2$, если $k_1 \le k_2$.

Нетрудно доказать, что $R_1 \gtrsim R_2 \Rightarrow R_1 \vdash R_2$ и $R_1 \succ R_2 \Rightarrow R_1 \vdash R_2$, где $\vdash$ – доказуемость в исчислении высказываний.

Таким образом, не менее общие (и более общее) высказывания логически сильнее. Кроме того, более общие правила проще, так как содержит меньшее число литер в посылке правила, поэтому отношение $\succ$ можно воспринимать как *отношение простоты* в смысле [8].

Определим множество предложений F, как множество высказываний, полученных из литер L замыканием относительно логических операций $\wedge, \vee$.

Вероятность на множестве предложений F определим как отображение $\mu : F \mapsto [0,1]$, удовлетворяющее следующим условиям [9]:

1. Если $\vdash \varphi$, то $\mu(\varphi) = 1$;
2. Если $\vdash \neg(\varphi \wedge \psi)$, то
   $\mu(\varphi \vee \psi) = \mu(\varphi) + \mu(\psi)$.

Определим условную вероятность правила $R = (a_1 \,\&\, ... \,\&\, a_k \Rightarrow c)$ как

$$\mu(R) = \mu(c / a_1 \,\&\, ... \,\&\, a_k) = \frac{\mu(a_1 \,\&\, ... \,\&\, a_k \,\&\, c)}{\mu(a_1 \,\&\, ... \,\&\, a_k)},$$

если $\mu(a_1 \,\&\, ... \,\&\, a_k) > 0$. Предполагается, что эмпирическая вероятность, определённая выше, в пределе дает вероятность μ. Множество всех правил из Pr, для которых условная вероятность определена, обозначим через Pr_0.

Вероятностным законом будем назы-

вать такое правило $R \in \Pr_0$, которое нельзя обобщить (логически усилить) не уменьшив его условную вероятность, т.е. для любого $R' \in \Pr_0$, если $R' \succ R$, то $\mu(R') < \mu(R)$.

Вероятностные законы – это наиболее общие, простые и логически сильные правила, среди правил, имеющих не большую условную вероятность. Обозначим множество всех вероятностных законов через PL.

Формальную модель нейрона определим как множество всех вероятностных законов $\Phi = \{R\}, R \in PL$, которые обнаруживает нейрон.

Отношение *вероятностного вывода* $R_1 \sqsubseteq R_2$, $R_1, R_2 \in PL$ определим как одновременное выполнение двух неравенств $R_1 \succsim R_2$ и $\mu(R_1) \le \mu(R_2)$. Если оба неравенства строгие, то отношение вероятностного вывода будем называть строгим *отношением вероятностного вывода*

$$R_1 \sqsubset R_2 \Leftrightarrow R_1 \succ R_2 \ \& \ \mu(R_1) < \mu(R_2).$$

Семантическим вероятностным выводом [10-12] будем называть максимальную (которую нельзя продолжить) последовательность вероятностных законов, находящихся в отношении строгого вероятностного вывода $R_1 \sqsubset R_2 \sqsubset ... \sqsubset R_k$. Последний вероятностный закон R_k в этом выводе будем называть *максимально специфическим*.

Теорема [10,13]**.** Предсказание по максимально специфическим правилам непротиворечиво.

Нами разработана программная система Discovery, реализующая семантический вероятностный вывод, которая успешно применялась для решения ряда прикладных задач [13-14].

Заключение

Полученная формальная модель, с одной стороны, формализует правило Хебба, а, с другой стороны, позволяет делать непротиворечивые предсказания.

Список литературы

1. Витяев Е.Е., Принципы работы мозга, содержащиеся в теории функциональных систем П.К. Анохина и теории эмоций П.В. Симонова // Нейроинформатика, 2008, том 3, № 1, стр. 25-78
2. Витяев Е.Е. Формальная модель работы мозга, основанная на принципе предсказания // Модели Когнитивных Процессов. (Вычислительные системы, 164), Новосибирск, 1998, стр. 3-61
3. Демин А.В., Витяев Е.Е. Логическая модель адаптивной системы управления. Нейроинформатика, 2008, том 3, № 1, стр. 79-107
4. Hempel, C. G. ‘Maximal Specificity and Lawlikeness in Probabilistic Explanation’, Philosophy of Science 35, 1968. – P. 16–33.
5. Гибсон Дж. Экологический подход к зрительному восприятию. М.: Прогресс, 1988. С. 462.
6. Hebb D.O. The organization of behavior. A neurophysiological theory. NY, 1949. 335 p.
7. Русинова Е.В. Пластические перестройки нейронной активности сенсомоторной коры во время выработки клеточного аналога условного рефлекса // Журн. высш. нервн. деят. им. И.П. Павлова. 1977. Т. 27. С. 941–948.
8. Kovalerchuk B.,Ya., Perlovsky L.I. Dynamic logic of phenomena and cognition // IJCNN, 2008, pp. 3530-3537.
9. Halpern J.Y. An analysis of first-order logics of probability // Artificial Intelligence, 46, 1990, pp. 311-350.
10. Vityaev E.E. The logic of prediction // Mathematical Logic in Asia 2005, Proceedings of the 9th Asian Logic Conference, eds. Goncharov S.S., Downey R. and Ono.H., August 16-19, Novosibirsk, Russia, World Scientific, 2006, pp. 263-276.
11. Vityaev E.E., Smerdov S.O. New definition of prediction without logical inference // Proceedings of the IASTED international conference on Computational Intelligence (CI 2009), ed. Kovalerchuk B., August 17–19, Honolulu, Hawaii, USA, pp. 48-54.
12. Смердов С.О., Витяев Е.Е. Синтез логики, вероятности и обучения: формализация предсказания // Сибирские Электронные Математические Известия. Т.6, Институт математики им.С.Л. Соболева СО РАН, 2009, стр. 340-365.
13. Витяев Е.Е. Извлечение знаний из данных. Компьютерное познание. Модели когнитивных процессов. Новосибирский гос. ун-т. Новосибирск, 2006. 293 с.
14. Kovalerchuk B.Ya., Vityaev E.E. Data mining in finance: advances in relational and hybrid methods. Kluwer Academic Publisher, 2000, pp.308