МНОГОКАНАЛЬНЫЕ МОДЕЛИ ЗРЕНИЯ

В.Е. Дубровский, А.В. Гарусев
факультет психологии МГУ имени М. В. Ломоносова
vicdubr@mail.ru, percept5@mail.ru

Согласно современным представлениям (Goris R. L. T. et al., 2013), зрительная система может рассматриваться как множество параллельно работающих подсистем или каналов, “настроенных” на обнаружение определенных элементов предъявляемого стимула. Каждый из каналов реагирует на определенный диапазон пространственно-временных яркостных профилей, представленных в стимуле. Начиная с пионерской работы (Campbell, Robson, 1968) получено множество как психофизических, так и психофизиологических доказательств, свидетельствующих в пользу такого подхода (Lee, 2011). Для измерения характеристик каналов и их формального описания разработаны эффективные методы (Логвиненко, 1985; Дубровский, 2009).

Пусть испытуемому предъявляются двумерные изображения, характеризуемые пространственно-временными яркостными функциями $s = s(x, y, t, \lambda)$, зависящими от двух координат, времени и длины волны. С формальной точки зрения каналы $r_i = \varphi_i(s)$ являются функционалами, заданными на множестве стимулов. Каждому стимулу s ставится в соответствие набор выходов N каналов r_i, т.е. вектор-столбец $r = (r_1, r_2, \cdots, r_N)^T$. По сути дела, такая психофизическая модель зрительного анализатора является формализованным вариантом концепции системы нейронов-детекторов (Фомин, Соколов, Вайткявичюс, 1979).

В соответствии с имеющимися экспериментальными данными допустимо считать, что в первом приближении каналы удовлетворяют условиям линейности. Для любой пары стимулов s_1, s_2 реакция канала на взвешенную сумму равна взвешенной сумме реакций на каждый из стимулов:

$$\varphi_i(\alpha_1 s_1 + \alpha_2 s_2) = \alpha_1 \varphi_i(s_1) + \alpha_2 \varphi_i(s_2)$$

При этом условии реакцию канала можно записать в виде интеграла

$$r_i = \varphi_i(s) = \iiint_{\Omega} h_i(x, y, t, \lambda) s(x, y, t, \lambda)\, dx\, dy\, dt\, d\lambda$$

где $h_i = h_i(x, y, t, \lambda)$ – весовая функция канала. Будем считать, что функции s и h_i являются элементами гильбертова пространства, в котором этот интеграл определяет скалярное произведение: $r_i = \langle h_i, s \rangle$

Линейная многоканальная модель зрительного анализатора характеризуется набором весовых функций h_i $(i = 1,2, \dots, N)$ и задает отображение пространства стимулов (в общем

случае – бесконечномерного) на конечномерное пространство каналов. Весовая функция линейного канала совпадает с рецептивным полем нейрона.

В литературе описано множество многоканальных моделей, различающихся выбором весовых функций. Подробный обзор ранних работ имеется в монографии (Логвиненко, 1985). Для простоты будем рассматривать одномерный случай, т.е. $h_i = h_i(x)$.

Первоначально широко обсуждалась гипотеза, предполагающая, что весовые функции каналов – синусы и косинусы:

$$h_m^c(x) = cos[2\pi f_m x]$$
$$h_m^s(x) = sin[2\pi f_m x]$$

В этом случае отклики каналов являются коэффициентами разложения в ряд Фурье, так что данная модель рассматривает зрительную систему как Фурье-анализатор. Однако быстро стало понятно, что такая модель обладает рядом существенных недостатков: Каждый из коэффициентов Фурье несет информацию об изображении в целом, не позволяя выделять отдельные детали. Синусоидальные весовые функции заданы на бесконечном интервале, что соответствует бесконечно широким рецептивным полям гипотетических нейронов. Между тем, рецептивные поля реальных нейронов должны быть пространственно локализованы.

Естественно, у исследователей возникло желание “подправить” набор весовых функций таким образом, чтобы избежать подобных недостатков. В одной из первых работ такого рода (S. Marčelja, 1980) было предложено использовать систему функций Габора

$$h_{m,n}^c(x) = e^{-\frac{(x-x_n)^2}{4\sigma^2}} cos[2\pi f_m(x - x_n)]$$

$$h_{m,n}^s(x) = e^{-\frac{(x-x_n)^2}{4\sigma^2}} sin[2\pi f_m(x - x_n)]$$

Эти функции одновременно локализованы в пространстве вокруг точки x_n и в частотной области вокруг частоты f_m. По форме они похожи на определяемые экспериментально рецептивные поля. Легко показать, что чем точнее произвольная весовая функция локализована в пространстве, тем шире должен быть ее спектр в частотной области (и наоборот). Функции Габора являются оптимальными в том смысле, что их эффективная ширина минимальна и в пространственной, и в частотной области. Теоретически, набор каналов с такими весовыми функциями хорошо подходит для выделения специфических признаков в реальных сценах.

Дальнейшее развитие идеи выбора оптимального набора весовых функций для зрительных каналов связано с концепцией вейвлетов (Добеши, 2001). Сам термин “вейвлеты” был введен в работе (Grossman, Morlet, 1984) для обозначения набора базисных функций, получаемых из некоторой исходной функции при помощи сдвигов и изменений масштаба.

Для построения системы вейвлетов выбирается материнский вейвлет $\psi(x)$ - функция, удовлетворяющая условию $\int \psi(x)\,dx = 0$. В качестве примера можно привести часто используемую “мексиканскую шляпу”:

$$\psi(x) = (1 - x^2)e^{-\frac{x^2}{2}}$$

Весовыми функциями служат вейвлеты

$$h_{a,b}(x) = |a|^{-1/2}\psi\left(\frac{x-b}{a}\right)$$

Здесь параметр масштаба a соответствует периоду $T = \frac{1}{f}$ в разложении Габора.

Если необходимо получить дискретный набор вейвлетов, то выбираются константы $a_0, b_0 > 1$, и вычисляются $a = a_0^m$ и $b = nb_0a$. При этом получается семейство

$$h_{m,n}(x) = |a_0|^{-m/2}\psi(a_0^{-m}x - nb_0)$$

От традиционного преобразования Фурье вейвлет-преобразование отличается тем, что коэффициенты разложения зависят не только от частот, но и от координат, которые рассматриваются как независимые переменные.

Форма вейвлетов напоминает рецептивные поля, что естественно привело к появлению моделей зрительного анализатора, использующих этот аппарат (Sierra-Vazquez, García-Pérez, 1995). С точки зрения моделирования системы реальных нейронов такое описание дает вполне адекватные результаты, позволяя рассматривать зрительный анализатор как некоторую систему эффективного кодирования изображений (Mallat, 1996). Обычно задаются тем или иным критерием оптимальности, который порождает систему весовых функций. Их профили оказываются похожими на реальные рецептивные поля, определяемые методами сенсорной физиологии. По мысли авторов это является решающим аргументом в пользу подобных теорий. Особенно плодотворным такой подход оказывается при создании “бионических” систем искусственного зрения (Martens, 2003).

Между тем, при использовании психофизических методов изучения зрительной системы с помощью канального подхода остается главный вопрос, который авторы подобных работ старательно обходят. При предъявлении стимула на экране испытуемый видит не набор признаков, а целостное изображение. На основании этого он принимает решение, отвечая на поставленный экспериментатором вопрос. Как же соотносится набор откликов системы каналов с тем, что реально видит испытуемый?

Вполне непротиворечивую концепцию предложил, развивая идеи А.Н. Леонтьева, А.Д. Логвиненко в своей монографии (Логвиненко, 1985). При помощи психофизических методик и специальной тренировки испытуемых в эксперименте можно получить доступ к сенсорной основе образа – некоторому субъективному изображению, которое описывается функцией

светлоты $\tilde{s} = \tilde{s}(x, y, t)$. Тогда работу нижнего уровня зрительной системы можно формально представить как сенсорный оператор $\mathcal{L}: s \mapsto \tilde{s}$, отображающий множество яркостных функций в множество функций светлоты. Наша задача - в рамках этого подхода связать концепции многоканальной модели и сенсорного оператора.

Но сначала необходимо сделать одно замечание. Вообще говоря, для того, чтобы анализировать и интерпретировать некоторый сигнал как изображение, совершенно не обязательно, чтобы он действительно был изображением (картинкой на экране монитора, фотографией и т.п.). К примеру, рассмотрим одну из часто встречающихся схем передачи и обработки изображений. Реальная сцена при помощи видеокамеры превращается в аналоговый электрический сигнал. Этот сигнал записывается видеомагнитофоном на магнитную ленту. Далее он считывается с ленты, усиливается, превращается в радиосигнал и излучается в эфир. Радиосигнал принимается антенной, снова усиливается, оцифровывается и вводится в компьютер. Таким образом, мы имеем дело с целым набором различных физических сигналов, однако нигде на этом пути не возникает реальная картинка, на которую можно было бы посмотреть. Это не мешает нам оценивать все эти устройства именно в терминах качества изображения: разрешающая способность, искажения, уровень помех и т.п. Дело в том, что на каждом этапе нам известен алгоритм пересчета физического сигнала в изображение. Так как мы знаем, что эти устройства предназначены для работы с изображениями, нам удобно их оценивать именно в этих терминах.

Следуя этому подходу, можно рассматривать вектор откликов набора каналов как некоторое закодированное изображение. Тогда естественно задаться вопросом, насколько хороша эта система кодирования, и как восстановить изображение. Напомним, что только наличие алгоритма декодирования позволяет говорить об изображении в этом случае.

Наиболее естественным образом эта задача решается в терминах линейных векторных пространств. Итак, имеется яркостная функция входного изображения s и набор весовых функций каналов $\{h_i\}_{i=1}^{N}$. Каждое из скалярных произведений $r_i = \langle h_i, s \rangle$ пропорционально проекции вектора s на один из векторов h_i. Если бы вектора $\{h_i\}_{i=1}^{N}$ образовывали ортогональный базис в подпространстве $\mathcal{H} = span\{h_i\}_{i=1}^{N}$, являющемся их линейной оболочкой (т.е. содержащем все возможные линейные комбинациями этих векторов), то задача решалась бы просто как восстановление вектора по набору его координат. Именно поэтому вейвлеты стараются выбрать так, чтобы они были ортогональными – это сильно упрощает процедуру анализа. Но реальные весовые функции каналов совершенно не обязаны удовлетворять условию ортогональности, и для них $\langle h_i, h_j \rangle \neq 0$ при $i \neq j$. Более того, эти вектора вообще могут не быть базисом в натянутом на них подпространстве, т.е. может оказаться так, что эти вектора линейно зависимы.

С формальной точки зрения задача выглядит следующим образом: Задано отображение $s \mapsto r$, причем само изображение s неизвестно, а доступны только проекционные коэффициенты $r_i = \langle h_i, s \rangle$. Требуется восстановить вектор по его скалярным произведениям с заданным набором векторов, т.е. найти такое изображение $\tilde{s}$, чтобы ему соответствовали эти же проекционные коэффициенты. Понятно, что вектор $\tilde{s}$ должен лежать в подпространстве, натянутом на вектора $\{h_i\}_{i=1}^N$, являясь их линейной комбинацией. Если отказаться от этого условия, то к $\tilde{s}$ можно добавить произвольный вектор $\widetilde{\Delta s}$, ортогональный всем векторам h_i, и это также будет решением: если $\langle h_i, \widetilde{\Delta s} \rangle = 0$, то $r_i = \langle h_i, \tilde{s} \rangle = \langle h_i, \tilde{s} + \widetilde{\Delta s} \rangle$.

Будем искать такие коэффициенты разложения α_j, чтобы $\tilde{s} = \sum_j \alpha_j h_j$ и $\forall i: r_i = \langle h_i, \tilde{s} \rangle$. Решение должно удовлетворять системе линейных уравнений:

$$r_i = \langle h_i, \tilde{s} \rangle = \langle h_i, \sum_j \alpha_j h_j \rangle = \sum_j \alpha_j \langle h_i, h_j \rangle$$

Матрица с элементами $G_{i,j} = \langle h_i, h_j \rangle$ называется матрицей Грама. Наша задача свелась к решению матричного уравнения $r = G\alpha$, откуда $G^{-1} r = \alpha$. Обозначив через $\gamma_{i,j}$ элементы матрицы G^{-1}, получаем формулу для вычисления коэффициентов разложения $\alpha_j = \sum_i \gamma_{i,j} r_i$.

Для того, чтобы решение существовало, матрица Грама должна быть невырожденной. Это возможно только если вектора $\{h_i\}_{i=1}^N$ не являются линейно зависимыми, т.е. образуют базис в натянутом на них подпространстве $\mathcal{H}$.

Пусть это условие выполняется. Зададим новый набор векторов $h_j^* = \sum_k \gamma_{j,k} h_k$ так, чтобы выполнялось условие

$$\langle h_i, h_j^* \rangle = \sum_k \gamma_{j,k} \langle h_i, h_k \rangle = \begin{cases} 0, i \neq j \\ 1, i = j \end{cases}$$

Иначе говоря, потребуем, чтобы каждый из векторов h_j^* был ортогонален всем векторам h_i, кроме вектора h_j. Набор векторов $\{h_j^*\}_{j=1}^N$ также является базисом. Базисы $\{h_i\}_{i=1}^N$ и $\{h_j^*\}_{j=1}^N$ называются биортогональными или двойственными. С использованием биортогонального базиса выражение для восстановленного изображения сильно упрощается:

$$\tilde{s} = \sum_j \alpha_j h_j = \sum_j \sum_i \gamma_{i,j} r_i \, h_j = \sum_i r_i h_i^*$$

или

$$\tilde{s} = \sum_i \langle h_i, s \rangle h_i^*$$

Если вектора h_i взаимно ортогональны, т.е. $\langle h_i, h_j \rangle = \begin{cases} 0, i \neq j \\ \|h_i\|^2, i = j \end{cases}$, матрица Грама становится диагональной. Обратная к ней матрица G^{-1} легко вычисляется – она также

диагональная с элементами $1/\|h_i\|^2$ на диагонали. Отсюда получается обычная формула для разложения по ортогональному базису

$$\tilde{s} = \sum_i \frac{1}{\|h_i\|^2} \langle h_i, s \rangle h_i$$

При этом выполняется обобщенное равенство Парсеваля:

$$\|\tilde{s}\|^2 = \langle \tilde{s}, \tilde{s} \rangle = \langle \sum_i \frac{1}{\|h_i\|^2} \langle h_i, s \rangle h_i \,, \sum_j \frac{1}{\|h_j\|^2} \langle h_j, s \rangle h_j \rangle$$

$$= \sum_i \sum_j \frac{1}{\|h_i\|^2} \frac{1}{\|h_j\|^2} \langle h_i, s \rangle \langle h_j, s \rangle \langle h_i, h_j \rangle = \sum_i \frac{1}{\|h_i\|^2} |\langle h_i, s \rangle|^2$$

Итак, если весовые функции каналов линейно независимы, можно вычислить двойственный им набор функций, являющихся элементами биортогонального базиса, и использовать их взвешенную сумму для восстановления изображения. Выходы каналов служат при этом весовыми коэффициентами.

Вообще говоря, условие линейной независимости весовых функций каналов противоречит требованию надежности кодирования изображения с помощью набора коэффициентов $r_i = \langle h_i, s \rangle$. Действительно, если рецептивные поля нейронов образуют базис, то исключение по какой-либо причине любого из таких нейронов ведет к невосполнимой потере части информации. Избыточность в такой системе отсутствует. Для того, чтобы распространить формализм на этот случай, введем понятие фрейма (Christensen, 2002;. Han Deguang et al, 2007).

Пусть предъявляемое изображение является элементом некоторого гильбертова пространства: $s \in \mathcal{S}$. Если существуют такие константы $A, B > 0$, что для любого $s \in \mathcal{S}$ выполняется условие

$$A\langle s, s \rangle \leq \sum_i |\langle h_i, s \rangle|^2 \leq B\langle s, s \rangle$$

то $\{h_i\}_{i=1}^N$ является *фреймом* (англ. *frame*, иногда переводят как *каркас*). Существование верхней границы гарантируется неравенством Коши-Шварца: $|\langle h_i, s \rangle|^2 \leq \langle h_i, h_i \rangle \langle s, s \rangle$, откуда $B = \sum_i \langle h_i, h_i \rangle = \sum_i \|h_i\|^2$. Для того, чтобы существовала нижняя граница, необходимо и достаточно, чтобы пространство $\mathcal{S}$ совпадало с линейной оболочкой системы весовых функций: $\mathcal{S} = \mathcal{H} = span\{h_i\}_{i=1}^N$. Практически это означает, что для предъявления испытуемому не используются стимульные изображения, которые он не может видеть ни при каких условиях.

Если верхняя и нижняя границы фрейма совпадают, то он называется жестким:

$$\sum_i |\langle h_i, s \rangle|^2 = A\langle s, s \rangle$$

Это - аналог равенства Парсеваля. Используя простые преобразования, получаем:

$$\sum_i |\langle h_i, s\rangle|^2 = \sum_i \langle h_i, s\rangle\langle h_i, s\rangle = \langle \sum_i \langle h_i, s\rangle h_i, s\rangle = A\langle s, s\rangle$$

Это условие должно выполняться для любого s, что возможно только если

$$s = \frac{1}{A}\sum_i \langle h_i, s\rangle h_i$$

Полученное выражение, решающее задачу восстановления изображения для жесткого фрейма, очень похоже на приведенную выше формулу разложения по ортогональному базису. Тем не менее, жесткий фрейм не обязательно является ортогональным базисом. Естественно задаться вопросом: какие дополнительные ограничения необходимо наложить, чтобы он превратился в ортогональный базис?

Выберем $s = h_k$. Тогда $A\langle h_k, h_k\rangle = \sum_i |\langle h_i, h_k\rangle|^2 = |\langle h_k, h_k\rangle|^2 + \sum_{i\neq k} |\langle h_i, h_k\rangle|^2$, или $\langle h_k, h_k\rangle(A - \langle h_k, h_k\rangle) = \sum_{i\neq k} |\langle h_i, h_k\rangle|^2$. Если $\langle h_k, h_k\rangle = \|h_k\|^2 > A$, то левая часть выражения отрицательна, так что равенство невозможно. При $\langle h_k, h_k\rangle = \|h_k\|^2 = A$ сумма в правой части должна обращаться в нуль, что возможно только если $\forall i \neq k: \langle h_i, h_k\rangle = 0$. Это означает, в силу произвольности выбора h_k, что $\{h_i\}_{i=1}^N$ - ортогональный базис.

Таким образом, для всех элементов жесткого фрейма должно выполняться условие $\|h_k\|^2 \leq A$, причем если $\forall k: \|h_k\|^2 = A$, жесткий фрейм является ортогональным базисом.

В общем случае для фрейма определен оператор $Fs = \sum_i \langle h_i, s\rangle h_i$. Можно показать, что обратный оператор F^{-1} существует, и может быть применен к обеим частям этого выражения, что приводит к формуле обращения:

$$s = F^{-1}\left(\sum_i \langle h_i, s\rangle h_i\right) = \sum_i \langle h_i, s\rangle F^{-1} h_i$$

Пусть $h_i^* = F^{-1}h_i$. Множество $\{h_i^*\}_{i=1}^N$ образует фрейм, двойственный к $\{h_i\}_{i=1}^N$. Использование двойственного фрейма позволяет представить формулу обращения в более наглядном виде:

$$s = \sum_i \langle h_i, s\rangle\, h_i^*$$

Итак, если набор весовых функций некоторого множества линейных каналов образует фрейм, то по набору выходов этих каналов можно однозначно восстановить предъявляемое изображение. Точнее говоря, находится проекция этого изображения на подпространство минимальной размерности, которому принадлежат все весовые функции. Таким образом решена общая задача о построении сенсорного оператора по множеству линейных каналов.

Предложенная модель намечает подходы сразу к нескольким проблемам. Во-первых, теория фреймов позволяет для каждой многоканальной модели оценить качество и

надежность кодирования изображений, и с этой точки зрения сравнить различные модели. Во-вторых, полученные соотношения дают возможность соотнести психофизические данные, полученные с использованием двух различных методов. С одной стороны, это пороговые эксперименты, связанные с идентификацией характеристик каналов, с другой - надпороговые эксперименты типа подравнивания контрастов, оценок качества предъявляемых изображений и их искажений и т.п.

Литература

Добеши И. Десять лекций по вейвлетам // М., РХД, 2001

Дубровский В.Е. Геометрический подход к задаче сенсорного различения // “Современная психофизика” Под ред. В.А. Барабанщикова Изд-во “Институт психологии РАН”, М., 2009, 110-144

Логвиненко А.Д. Чувственные основы восприятия пространства // М., МГУ, 1985

Фомин С. В., Соколов Е. Н., Вайткявичюс Г. Г. Искусственные органы чувств // М., Наука, 1979

Campbell F.W., Robson J.G. Application of Fourier analysis to the visibility of gratings // J. Physiol., 1968; 197(3), 551–566

Christensen O. An introduction to frames and Riesz bases // Springer, 2002

Goris R. L. T. et al. A neural population model for visual pattern detection // Psychol. Rev., 2013, 120(3), 472-496

Grossmann A., Morlet J. Decomposition of Hardy Functions into Square Integrable Wavelets of Constant Shape // SIAM J. Math. Anal., 1984, 15(4), 723–736

Han Deguang, Kornelson K., Larson D., Weber E. Frames for undergraduates // Student Mathematical Library Vol. 40 // Amer. Math. Soc., 2007

Lee B. B. Visual pathways and psychophysical channels in the primate // J. Physiol., 2011, 589(1), 41-47

Mallat S. Wavelets for a vision // IEEE Proc., 1996, 84(4), 604-614

Marčelja S. Mathematical description of the responses of simple cortical cells //JOSA, 1980, 70(11), 1297-1300

Martens J.B. Image Technology Design: A Perceptual Approach // The Springer International Series in Engineering and Computer Science // Springer, 2003

Sierra-Vazquez V., García-Pérez M. A. Psychophysical 1-D wavelet analysis and the appearance of visual contrast illusions // IEEE Trans. on Systems, Man and Cybernetics, 1995, 25(10), 1424-1433

Дубровский Виктор Ефимович, научный сотрудник, факультет психологии МГУ имени М. В. Ломоносова, vicdubr@mail.ru, 8-916-607-47-33

Гарусев Александр Владимирович, научный сотрудник, факультет психологии МГУ имени М. В. Ломоносова, percept5@mail.ru, 8-916-642-62-74