

Нейросетевой тест на вхождение в сетевое сообщество



Представление специфической информации в мозге человека в виде иерархии словарей образов событий разной частоты встречаемости и разных модальностей является моделью мира человека, состоящей из трех компонентов — индивидуального многомодального правополушарного (у правшей) и двух социализированных левополушарных — многомодального и лингвистического, которые (все) в какой-то степени изоморфны друг другу. Наверное, менталитет человека определяется его моделью мира. Сравнивая модели мира (их лингвистические компоненты) отдельных людей, возможно, удастся понять степени близости их менталитетов. Поэтому можно попытаться интерпретировать менталитет авторов текстов на основе анализа их текстов. Сравнивая тексты членов разных «сетевистских» сообществ (близость внутри, различие между ними), возможно, удастся выяснить принципы объединения людей в этих сообществах.

УДК 642.825.8

Ключевые слова

«Сетевистское сообщество», обработка информации в мозге, ламели гиппокампа, лингвистический компонент модели мира, семантическая сеть.

Автор

Харламов Александр Александрович — доктор технических наук.

В последнее время в социально-политических публикациях поднимается вопрос о том, что общество из рыночного превращается в сетевое: все больше социальных сообществ организуется по типу сети. Наиболее характерный пример — социальные сети. Стремление людей объединиться по «сетевистскому» типу обусловлено их ментальной близостью. Попробуем понять, что же лежит в основе такой близости. Очевидно, эта близость определяется одинаковостью взглядов, вследствие чего членам «сетевистского» сообщества комфортно вместе. Возможно, такая близость точек зрения на сущее есть следствие схожести их моделей мира. Попробуем это объяснить на основе понимания процессов обработки информации, протекающих в мозге человека.

Мозг человека является естественной нейронной сетью. В отличие от искусственных нейронных сетей эта естественная нейронная сеть хорошо структурирована. Именно поэтому (и еще потому, что нейронов в ней очень много) она детально отражает мир в терминах сенсорных и эффекторных модальностей, присущих нервной системе. Архитектура этой естественной нейронной сети очень сложна. Но отдельные ее модули представляют собой повторяющиеся структуры, функциональность которых сравнительно несложно интерпретируется. Понимание этой функциональности дает возможность представить, что делается в мозге со специфической информацией, поступающей на его входы, и как она там хранится. Такое пред-



ставление информации фактически опосредует моделирование мира, то есть является моделью мира.

Одна из частей модели мира человека — лингвистическая. Действительно, эта модель состоит из трех взаимосвязанных частей: индивидуального (правополушарного у правой) и социализированного (левополушарного) многомодальных компонентов, а также ее (модели мира) лингвистического компонента. Поскольку язык опосредует мышление человека, без натяжки можно сказать, что лингвистическое представление модели мира изоморфно всей модели, а анализ текстов как порождений этой модели тождествен анализу самой модели. Попробуем понять на основе анализа текстов, в чем же заключается близость точек зрения, моделей мира, менталитета и т.п. Сначала выясним, что же это такое — модель мира, как она формируется и где представлена. Наверное, мы не сделаем открытия, если скажем, что модель мира формируется в мозге человека. Рассмотрим этот субстрат и этот процесс.

Нейронная сеть в мозге

Мозг человека представляет собой сложную неоднородную естественную нейронную сеть, разобраться в которой физиологам до сих пор не удалось из-за отсутствия технологий анализа совместной активности больших цепочек нейронов [1]. Однако если иметь некоторое представление о совместной обработке информации в нескольких отдельных структурах мозга, в первую очередь в колонках коры и в ламелях гиппокампа [2], представляется реальным интерпретировать обработку и хранение информации в мозге, что позволяет понять, как устроена модель мира человека. Рассмотрим, как устроены и как работают колонка коры и ламель гиппокампа, так как основной объем обработки специфической ин-

формации, поступающей в мозг и исходящей из мозга, осуществляется именно в них.

Кора

Кора представляет собой тонкий слой нейронов на поверхности полушарий большого мозга человека. Вертикальный срез коры состоит из морфологически различных нейронов, собранных в слои. Наибольший интерес с точки зрения обработки информации представляют так называемые пирамидные нейроны третьего слоя коры. Они интересны тем, что их реакция на различные последовательности приходящих на них сигналов избирательна. Таким образом, их можно избирательно возбуждать, подавая на них те или иные информационные последовательности (эти последовательности совпадают по структуре с так называемым адресом нейрона) [2].

Пирамидные нейроны собраны в так называемые колонки (компактные структуры, формируемые пирамидными нейронами 3-го слоя коры, объединенными общим информационным входом и общим управляющим входом), которые собираются в гиперколонки. Свойство пирамидных нейронов возбуждаться разными информационными последовательностями называется ассоциативным обращением. Такие структуры из пирамидных нейронов моделируют фрагменты многомерного пространства — многомерного гиперкуба. Координаты точек такого фрагмента пространства, этой гиперколонкой смоделированного, суть те самые входные последовательности, которые являются характерными для разных пирамидных нейронов гиперколонки.

Траектория в многомерном пространстве

Рассмотрим, как формируются последовательности, кото-

рые являются входными для колонок коры. Любая информация поступает через сенсоры в мозг как некоторая информационная последовательность. Речевой сигнал — это одномерная последовательность. Видео-ряд — двумерная последовательность. То или иное также можно сказать и об информации других модальностей. Преобразованная в периферии сенсорных органов информация сохраняет форму последовательностей во времени. Кодировка этой информации характерна для кон-

Стремление людей объединиться по «сетевистскому» типу обусловлено их ментальной близостью.



кретных сенсоров. Практически все сенсоры мозга человека представляют собой матрицы, что приводит к появлению множества параллельно идущих последовательностей. Эти последовательности распределяются между гиперколонками коры топологически, сохраняя распределение модальностей по поверхности полушарий мозга. Ассоциативность обращения (свой адреса к своей гиперколонке) приводит к тому, что близкая по форме информация адресуется в близкие колонки. Каждая такая информационная последовательность, поступающая на свою гиперколонку, приводит к последовательному возбуждению тех или иных нейронов, ее составляющих, в зависимости от того или иного ее фрагмента, который является адресом того или иного нейрона. Напомним, что в совокупности нейроны гиперколонки моделируют некоторый фрагмент многомерного пространства, в котором в ответ на входную информационную последовательность возникает последовательность возбужденных нейронов — так называемая траектория.

Формирование словарей

Ассоциативность обращения входной последовательности к нейронам гиперколонки порождает интересное свойство преобразования информации в многомерном пространстве. Как только во входной последовательности появляется повторяющийся фрагмент, траектория в многомерном пространстве попадает в участок, который был ранее пройден при наличии аналогичного фрагмента во входной последовательности.

Так как все входные сенсорные и выходные эффекторные информационные последователь-

Условно модель мира можно представить как совокупность слов словарей — образов событий разных уровней и разных модальностей.

ности являются квазитекстами (они внутренне структурированы), они содержат повторяющиеся фрагменты — элементы разных уровней представления информации. Так, в речевой последовательности эти повторяющиеся фрагменты — фонемы, морфемы, слоги, слова и т.д. В результате отображения в многомерное пространство любая входная последовательность порождает траекторию, формирующую пучности в местах прохождения участков, соответствующих повторяющимся фрагментам последовательности. Убирая редко встречающиеся участки траектории (например, пороговым преобразованием), мы автоматически выявляем эти повторяющиеся фрагменты. В гиперколонке формируется словарь событий заданного уровня (заданной частоты встречаемости).

В процессе обработки поступающей постоянно на вход коры информации постепенно запол-

няются гиперколонки разных областей коры. При этом формируются словари событий разных уровней различных модальностей. Так, в слуховой коре, в ее первичной проекционной зоне (зона Вернике), отвечающей за восприятие речи, формируется словарь аллофонов, в окружающих близлежащих полях и подполях формируются словари флективных и корневых морфем, а также суффиксов, префиксов и т.д.

Таким образом формируется многомодальная модель мира — множество словарей разных модальностей разных уровней, слова которых связаны друг с другом по ассоциации: близкое с близким. То же верно и для областей коры, формирующих образы событий других модальностей, в первую очередь зрительной. Есть также отдел коры — теменные доли, — в котором информация достаточно высоких уровней представления различных модальностей объединяется и далее обрабатывается совместно (также с формированием словарей образов событий, но более высоких уровней абстракции).

Гиппокамп

Другая структура мозга (гиппокамп) формирует из этих отображений событий внешнего мира — слов словарей, хранящихся в колонках коры, — комплексные представления, характеризующие целые ситуации. Если растянуть одну из симметричных частей гиппокампа вдоль длинной оси и разрезать ее поперек, на срезе мы увидим структуру (ее называют ламелью), которая и хранит информацию об объединении образов отдельных событий — слов словарей колонок коры — в целую ситуацию. Пирамидные нейроны поля CA₃ ламели гиппокампа, моделируемые искусственной нейронной сетью Хопфилда [3], которая в этом случае реализует ассоциативную память,

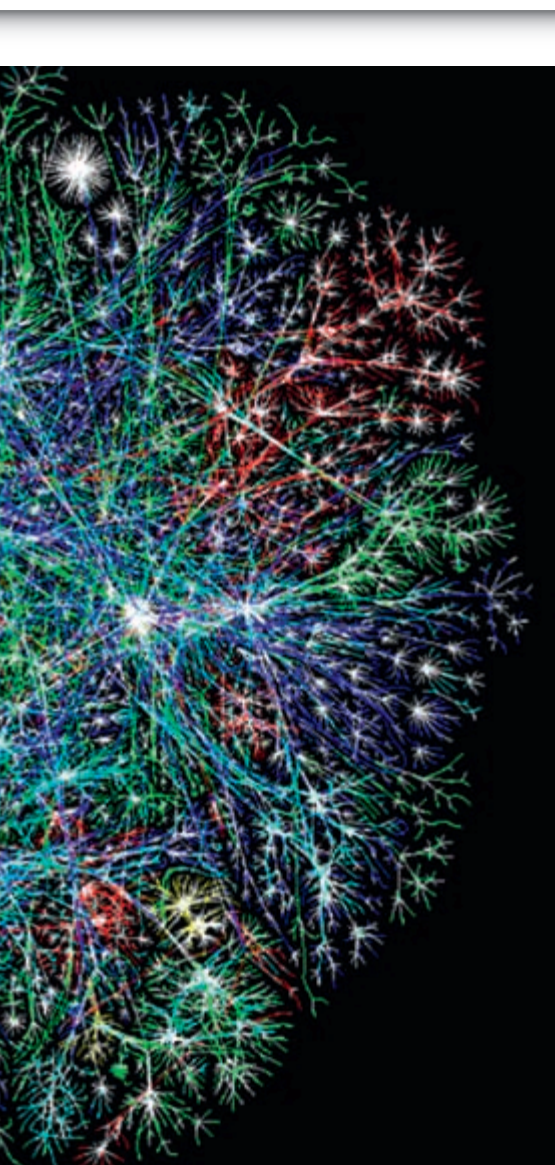
хранят представление о связях всех событий — слов различных словарей различных модальностей, — относящихся к ситуации. Таких ламелей в гиппокампе много, и каждая хранит информацию о своей ситуации.

При формировании моделей ситуаций связи между образами событий записываются в ассоциативную память. При распознавании входной ситуации входная информация по ассоциации адресуется к одной или нескольким ламелям, хранящим информацию о ситуациях, в большей или меньшей степе-





ни похожих на входную. Из них выбирается наиболее похожая. В зависимости от режима либо принимается решение о степени близости, либо входная информация дописывается в ламель, доформируя модель ситуации. Таким образом, в ламелях накапливается информация о ситуациях, постепенно усложняясь, детализируясь, так же как и в коре формирующая сгущения в местах, наиболее характерных для конкретных ситуаций. Благодаря гиппокампу мы можем оценить степень новизны приходящей информации.



Модель мира

Таким образом, в двух основных (для обработки специфической информации) органах мозга — коре и гиппокампе — в течение жизни человека формируется представление о мире в виде словарей образов событий разных уровней (разной сложности, разной частоты встречаемости), разных модальностей (сенсорных и эффекторных) в колонках коры и объединенных в представлении ситуаций — в ламелях гиппокампа. Это интегральное представление о мире — модель мира — имеет более крупную структуру. Она (модель мира) делится на три части: две из них находятся в левом полушарии (у правшей), а одна — в правом. В правом полушарии формируется индивидуальная модель мира — там хранятся образы всех индивидуальных событий, которые происходят в течение жизни человека. В левом полушарии (у правшей) имеется многомодальная схематическая (социализированная — построенная под воздействием социума) модель мира и лингвистическая модель мира, которая, собственно, и порождает схематическое многомодальное представление, сегментируя мир терминами, порожденными социумом.

Условно модель мира можно представить как совокупность слов словарей — образов событий разных уровней и разных модальностей (то же и в гиппокампе с образами ситуаций), но фактически раздельных слов в этих представлениях не наблюдается. Все это информационное многообразие, порожденное траекториями в многомерном пространстве, является неразрывным целым — сложным графом, слова представляются пучностями на этом графе, причем отдельное слово — совсем не обязательно цепочка точек пространства, а ветвистый подграф, каждая веточка которого характеризует свои особенности образа.

Лингвистический компонент модели мира

Как и остальные части модели мира человека, ее лингвистический компонент представляет собой множество словарей, но уже лингвистических событий, причем различных уровней традиционных представлений лингвистической информации: морфологического, лексического, синтаксического, семантического и прагматического. Они организованы и формируются так же, как и словари многомодальных компонентов модели, формируются и хранятся также в колонках коры, но в отличие от многомодальных словарей сравнительно легко интерпретируются, так как содержащаяся в них информация опосредована текстами (содержится в текстах, извлекается из текстов).

Поскольку все три компонента модели мира человека — лингвистический, многомодальный схематический и многомодальный индивидуальный — связаны по ассоциации горизонтальными связями (одинаковые события в разных модальностях с одинаковыми событиями), все три компонента оказываются изоморфны друг другу и всей модели в целом. Другими словами, анализируя тексты, мы можем сравнительно легко понять, как устроена обработка информации и ее хранение в лингвистическом компоненте модели мира, а следовательно, и в остальных двух и в их совокупности — во всей модели.

Обработка информации в лингвистическом компоненте модели мира сравнительно хорошо известна [2]. Эта обработка моделируется в достаточно хорошо распространенном программном продукте для автоматической смысловой обработки информации *TextAnalyst* [4, 5]. Обработка текстовой информации включает в себя формиро-



вание словарей лингвистических событий всех необходимых уровней, включая словарь флективных морфем (окончаний слов) на нижнем уровне, словарь корневых основ на лексическом уровне, словарь сочетаемости пар слов (по смыслу) на семантическом уровне. Предыдущие уровни обработки текстовой информации необходимы для формирования словаря пар слов. Наличие устойчивых комбинаций слов, обусловленных их семантикой, а также использование ассоциативности

ный университет им. Н.И. Лобачевского в 2005 г., представленный на интернет-сайте <http://www.unn.ru/rus/priem.htm>, и два других текста — устав Нижегородского государственного университета им. Н.И. Лобачевского и статья «Математический тривиум» из журнала «Успехи математических наук» (текст 2 и текст 3) из той же и из другой предметных областей. Степени пересечений семантических сетей исходного текста с сетями двух других текстов представлены в *табл. 1*.

Можно идти и от обратного: у людей, принадлежащих к одному «сетевистскому» сообществу, и модели мира должны быть близки. Так, у «сетевистского» сообщества синергетиков, статьи которых представлены на сайте [//spkurdyumov.ru](http://spkurdyumov.ru), должны быть похожи модели предметной области «Синергетика», то есть описывающие эту предметную область корпусы текстов должны порождать близкие по структуре семантические сети. Такие же корпусы текстов, относящихся к предметной области, например «Речевые технологии» (см. сайт журнала «Речевые технологии» speechtechnology.ru), должны порождать семантические сети, отличные от первых.

Наверное, менталитет человека определяется его моделью мира. Сравнивая модели мира отдельных людей, возможно, удастся понять степень близости их менталитетов.

обращения к информации, позволяет построить семантическую сеть, характеризующую целый текст: пара за парой, где второе слово первой пары является первым словом второй пары. Именно такая сеть является структурой, которая представляет смысл текста (так же можно обрабатывать и корпусы текстов).

Сравнение текстов по смыслу

Такие сети можно использовать для разных целей. В том числе их можно сравнивать (они представляют собой метрические пространства, поскольку ключевые понятия, их составляющие, в процессе формирования сети взвешиваются). А значит, можно сравнивать по смыслу им соответствующие тексты: чем больше степень пересечения сетей, тем больше степень смыслового подобия текстов.

Рассмотрим функцию сравнения семантических сетей на примере сравнении трех текстов, которые можно скачать в Сети. Пусть имеются некий русскоязычный текст (текст 1), например текст о правилах приема в Нижегородский государствен-

ный университет им. Н.И. Лобачевского в 2005 г., представленный на интернет-сайте <http://www.unn.ru/rus/priem.htm>, и два других текста — устав Нижегородского государственного университета им. Н.И. Лобачевского и статья «Математический тривиум» из журнала «Успехи математических наук» (текст 2 и текст 3) из той же и из другой предметных областей. Степени пересечений семантических сетей исходного текста с сетями двух других текстов представлены в *табл. 1*.

Ментальная близость. Отнесенность к «сетевистскому» сообществу

Как мы замечали выше, лингвистический компонент модели мира человека изоморфен всей модели, то есть, анализируя корпус текстов, порожденный индивидуумом, можно говорить о ментальной природе человека. Сравнивая по смыслу корпусы текстов, порожденные одним человеком, с корпусами текстов, порожденными другим человеком, можно пытаться соотнести модели мира двух разных людей (и большего количества тоже).

Попробуем сравнивать тексты, принадлежащие корпусам, описывающим указанные предметные области. Сначала выясним, как тексты одной предметной области соотносятся в этом смысле с моделью предметной области. Моделью предметной области будем считать множество текстов, ее описывающих. Построим модель предметной области на основе множества обучающей выборки текстов. Далее сравниваем тексты тестирующей выборки с моделью предметной области.

В качестве обучающей выборки для предметной области «Синергетика» возьмем восемь текстов с сайта «Синергетика», а для предметной области «Речевые технологии» — статьи 4-го номера журнала «Речевые технологии» за 2012 г. В качестве тестирующих выборок возьмем семь статей с сайта «Синергетика» (тексты 11–17) и четыре статьи 1-го но-

Таблица 1

| Степени пересечений семантических сетей исходного текста с сетями других текстов | | | |
|--|----------|----------|----------|
| Исходный текст | Текст 1 | Текст 2 | Текст 3 |
| Текст 1 | 100,0000 | 8,3533 | 2,9198 |
| Текст 2 | 8,3533 | 100,0000 | 1,1539 |
| Текст 3 | 2,9198 | 1,1539 | 100,0000 |

Таблица 2

| Степень близости тестовых текстов из предметной области «Синергетика» к модели предметной области «Синергетика» | | | | | | | |
|---|----------|----------|----------|----------|----------|----------|----------|
| | Текст 11 | Текст 12 | Текст 13 | Текст 14 | Текст 15 | Текст 16 | Текст 17 |
| Степень близости | 4,796 | 7,489 | 26,637 | 11,289 | 35,942 | 6,126 | 1,651 |

мера журнала «Речевые технологии» за 2013 г. (тексты 21–24).

Относительные расстояния текстов тестирующих выборок до моделей предметных областей приведены в *табл. 2* и *3*.


Далее вычислим относительные расстояния тестовых текстов предметной области «Синергетика» с моделью предметной области «Речевые технологии» (*табл. 4*).

Полученные расстояния свидетельствуют, что тексты одной предметной области находятся ближе друг к другу по смыслу, чем тексты разных предметных областей, что в общем-то тривиально, но иллюстрирует приведенные построения. Необходимо, правда, помнить, что качество анализа степени смысловой близости зависит и от качества текста, а также от его объема.

Рассмотрев существующие представления о строении двух основных (для обработки специфической информации) органов большого мозга человека — коры и гиппокампа, удалось проинтерпретировать обработку специфической информации в мозге человека с формированием модели мира в виде иерархии словарей образов событий разной частоты встречаемости и разных модальностей. Наверное, менталитет человека определяется его

Таблица 3

| Степень близости тестовых текстов из предметной области «Речевые технологии» к модели предметной области «Речевые технологии» | | | | |
|---|----------|----------|----------|----------|
| | Текст 21 | Текст 22 | Текст 23 | Текст 24 |
| Степень близости | 1,453 | 1,371 | 8,475 | 2,291 |

моделью мира. Сравнивая модели мира отдельных людей, возможно, удастся понять степень близости их менталитетов. Поскольку лингвистический компонент модели мира изоморфен всей модели, можно попытаться интерпретировать менталитет авторов текстов на основе анализа их текстов. Сравнивая тексты членов разных «сетевистских» сообществ (близость внутри, различие между ними), возможно, удастся выяснить принципы объединения людей в этих сообществах. 

ПЭС 14097/29.07.2014

Литература

1. Черч Дж., Юсте Р. Новая эра в изучении мозга // В мире науки. 2014. № 5. С. 6–12.
2. Харламов А.А. Нейросетевая технология представления и обработки информации (естественное представление знаний). М.: Радиотехника, 2006. 89 с.
3. Brown T.H., Zador A.M. Hippocampus. In G.M.Shepherd (Ed.), The synaptic organisation of the brain. NewYork, Oxford: Oxford University-Press, 1990. P. 346–388.
4. <http://www.analyst.ru>.
5. Sullivan Dan Document Warehousing and Textmining. NY; Wiley publishing house, 2001, ISBN 0471399590. 542 p.



Таблица 4

| Степень близости тестовых текстов из предметной области «Синергетика» к модели предметной области «Речевые технологии» | | | | | | | |
|--|----------|----------|----------|----------|----------|----------|----------|
| | Текст 11 | Текст 12 | Текст 13 | Текст 14 | Текст 15 | Текст 16 | Текст 17 |
| Степень близости | 0,000 | 0,000 | 3,003 | 0,144 | 2,923 | 0,727 | 0,000 |