ИНФОРМАЦИЯ И МОДЕЛИ[1]

Виктор Неделько

***Аннотация**: В работе обсуждается возможность формализации базовых понятий интеллектуальной деятельности, таких как информация, смысл (семантика), интеллект. Основная трудность этой задачи в том, чтобы достичь достаточно строгой математической формализации, сохранив содержательный смысл исходных понятий. Для достижения данной цели в работе используется подход, основанный на теории моделей, который позволяет формализовать данные понятия, определив их в конечном счете как некоторые множества. Исследуется возможность использования теории моделей в задаче понимания естественного языка и для оценивания семантической близости высказываний.*

***Ключевые слова**: информация, теория моделей, искусственный интеллект, экспертные системы, обработка естественного языка, семантика, высказывания на естественном языке.*

***ACM Classification Keywords**: I.2.0 Artificial intelligence – Philosophical foundations, I.2.7 Natural Language Processing, H.1 Information Systems – Models and principles.*

***Conference**: The paper is selected from XIVth International Conference "Knowledge-Dialogue-Solution" KDS 2008, Varna, Bulgaria, June-July 2008*

Введение

Целью данной работы является исследование возможности формализации основных понятий, связанных с интеллектуальной деятельностью, в частности, понятий информации, смысла высказывания, естественного и искусственного интеллекта.

Подобные формализации, вообще говоря, достаточно давно известны, например, теория информации, машины Тьюринга, однако они несут, как правило, существенно более узкий смысл, чем тот, который вкладывается в соответствующие понятия вне этих теорий. В данной работе представлена попытка избежать обеднения таких понятий при формализации, возможно, за счет отступления от математической строгости, но добиться, чтобы определения имели математический смысл, а именно, так или иначе определяли некоторое множество объектов с заданными свойствами.

Работа в значительной мере носит философский характер, но, пожалуй, в меньшей, чем может показаться на первый взгляд. Так многие моменты, которые могут показаться изложением идеалистических философских концепций, являются, как правило, не более чем частичной математической формализацией предметной области.

Базовые понятия

Базовыми понятиями для описания интеллектуальной деятельности являются понятия субъекта и объекта, которые мы можем в контексте данной работы отождествить с первичными философскими категориями «Я» и «не Я». «Я» – есть в первую очередь сознание, которое можно представить как совокупность ощущений (образов). Ощущения здесь – не только сигналы от органов чувств, но и мысленные образы (фактически такие же сигналы, только сымитированные мозгом).

Рассмотрим множество субъектов. Мироощущение каждого индивидуально, но для них имеет место некоторая эквивалентность.

Предположим, два субъекта сидят за столом и смотрят на книгу в зеленом переплете. Возможно, что их восприятие существенно различается, например, может оказаться, что зеленый цвет одним воспринимается так же, как другим красный. Однако установить это различие можно только, если произвести обмен сознаниями. Поскольку замена сознания возможна лишь умозрительно, для этих лиц

[1] Работа выполнена при поддержке РФФИ, грант № 07-01-00331-а.

возможно договориться, что они оба наблюдают зеленую книгу, и это ни в чем не приведет к противоречиям.

Таким образом, есть имеет место «изоморфизм» восприятия. Все множество ощущений можно факторизовать по отношению эквивалентности, иными словами, сопоставить каждому классу эквивалентных ощущений (образов) элемент некоторого множества. Полученное фактор-пространство назовем множеством моделей или универсумом.

Часть универсума, которая является общей для большинства субъектов, назовем реальностью.

Заметим, что такое определение реальности вовсе не подразумевает идеалистического подхода и не зависит от философских концепций. Философские аспекты вопроса более подробно обсуждаются в последнем разделе.

Связь с теорией моделей

Если говорить неформально, семантика высказывания — это те мысли, которые хотел передать автор. Однако мысли — сущность, недоступная для наблюдения, и для их описания нужны вспомогательные средства. В идеале, было бы сконструировать некоторый математический объект, который можно было бы сопоставить мыслеобразам и использоваться в качестве представления последних.

Одно из таких средств давно известно и широко используется – это модели, или идеализированные представления реальности. В качестве примеров можно привести понятия физического тела или материальной точки из физики. Наиболее строго этот подход развит в теории моделей. Теория моделей [Chang, Keisler 1973] – раздел математической логики, в котором вводится понятие модели как объекта, на котором можно определить истинность логического высказывания (предложения, формулы).

Как известно, формулами в логике высказываний являются сконструированные по определенным правилам последовательности из высказывательных символов $A, B, C, \ldots$, логических связок $\wedge, \vee, \neg$ и скобок $(,)$. Моделями в этом случае являются любые подмножества высказывательных символов. Если мощность множества высказывательных символов равна *n*, то число моделей составит 2^n. При этом количество классов эквивалентных, то есть не различимых на этих моделях, высказываний равно 2^{2^n}.

Логика высказываний является наиболее простой иллюстрацией применения теории моделей. В логике предикатов построение моделей уже существенно сложнее.

В данной работе нас интересует возможность использования подобного подхода при работе с естественным языком. При этом не ставится целью достичь такого же уровня формализации.

Выбор множества моделей определяется предметной областью. Например, если речь идет о зрительных образах и сценах, то моделями естественно выбрать комбинации геометрических фигур. Открытой является проблема построения универсального набора моделей.

Вполне естественно смыслом высказывания назвать множество моделей, на которых оно истинно. Однако для естественного языка истинность не всегда определяется однозначно.

Информация

В логике высказывание либо истинно на модели, либо ложно. Для естественного языка строгая истинность редко имеет место.

Для формализации неопределенности, присущей естественному языку, возможно использовать нечеткую логику или вероятностный формализм. Более адекватным представляется вероятностный подход, поскольку говорящий, как правило, вкладывает в высказывание вполне определенный смысл, который нам неизвестен. Нечеткие множества были бы адекватны, если бы в высказывание вкладывалось несколько смыслов одновременно, с разной степенью выраженности. Аргументом в пользу использования вероятности будет тот факт, что вероятность является фундаментальным понятием, используемым для описания физического мира (в квантовой механике, статистической физике).

Чтобы можно было определить вероятностную меру, на множестве моделей формально должна быть задана σ–алгебра событий. Это требование не создает проблем, поскольку человек различает лишь конечное число градаций чего бы то ни было, поэтому множество моделей всегда можно взять конечным.

Под информацией будем понимать распределение вероятностей на множестве моделей. Смыслом высказывания назовем его функцию правдоподобия на множестве моделей. Используя функцию правдоподобия, можно от априорного распределения перейти к апостериорному (см. раздел о согласовании экспертных высказываний).

В случае, когда высказывание является логической формулой, а множество всех моделей конечно, мера сходства высказываний может быть [Викентьев, 2004] введена как отношение числа моделей, на которых высказывания либо одновременно истинны, либо одновременно ложны, к общему числу моделей.

В общем случае сходство высказываний определяется через сходство их функций правдоподобия. Последнее может вводиться на основе известных способов задания расстояний на распределениях. Пожалуй, наиболее известным из таких расстояний является энтропийная метрика Кульбака. Однако данное расстояние не учитывает, что на самих моделях может быть определено понятие близости, поэтому более подходящей представляется, например, транспортная метрика Монжа-Канторовича.

Интеллект

Одним из признанных видов интеллектуальной деятельности является логический вывод. Поскольку мы распространяем логические термины на всю деятельность сознания, естественно будет определить интеллект путем обобщения понятия логического вывода.

Назовем интеллектом способность интерпретировать высказывания на естественном языке в универсуме человеческого восприятия и выполнять логический вывод. Напомним, что для этого не обязательно обладать человеческим восприятием, достаточно оперировать моделями некоторого универсума, эквивалентного человеческому, поэтому данное определение подходит как для естественного, так и для искусственного интеллекта.

Однако человек оперирует не формальным языком, а образами (которые могут иметь разный уровень абстрагированности), то есть фактически моделями. Известно, что логический вывод можно осуществлять не только путем эквивалентных преобразований высказываний в соответствии с заданными правилами, но и путем установления эквивалентности высказываний на множестве моделей. Именно второй способ, видимо, и присущ естественному интеллекту, и его стоит использовать при моделировании последнего.

Воля и творчество

В известном тесте Тьюринга искусственный интеллект будет признан состоятельным, если в общении он будет неотличим от человеческого.

Но кроме способности поддерживать беседу, человеческий интеллект характеризуют творческие способности.

Очевидно, что творчество – наиболее трудноформализуемая составляющая интеллектуальной деятельности, даже если под формализацией понимать лишь строгое определение. Вполне возможно предположить, что творчество является лишь эффектом проявления определенных свойств используемых человеком моделей. Так, например, интуицию можно объяснить подсознательным манипулированием моделью предметной области, аналогично тому, как компьютер просчитывает динамические модели объекта. Однако возможно существование еще одной составляющей творчества, которую будем называть волей. Заметим, что в данном разделе воля определяется как некоторая логическая возможность, и не обсуждается вопрос доказательства ее реального существования.

Философское исследование понятия воли обычно проводится в отношении ее свободы [Шопенгауэр]. В определении, которое здесь будет предложено, воля и свобода фактически являются синонимами.

Сначала определим волю для элементарных частиц. Современная физическая теория постулирует невозможность точного предсказания в пределах области квантовой неопределенности. Определим волю как нечто, что конкретизирует исход в пределах области определения волновой функции.

Это же определение годится и для воли человека. Для пояснения рассмотрим пример.

Предположим, что мы усилитель теплового шума вмонтировали в робота, который этим шумом и управляется: каждую секунду смотрится значение напряжения и, если оно больше ноля, то робот

поворачивает направо, в противном случае он поворачивает налево. В промежутках между поворотами робот двигается в текущем направлении с некоторой скоростью.

Согласно постулату квантовой механики траектория движения этого робота будет в принципе непредсказуемой, то есть никакие дальнейшие открытия физических законов не дадут ключа к предсказанию его действий, которые абсолютно случайны, то есть не зависят ни от чего из того, что нам известно или дано в ощущение.

Возникает вопрос, не могут ли сходные эффекты иметь отношение к поступкам человека? Современные знания по нейроанатомии и нейрофизиологии не исключают возможность того, что решения субъекта существенно зависят от чисто квантовых эффектов в микромире нервных клеток. Такая возможность кажется вполне правдоподобной, при этом она даёт физическое содержание понятию воля.

В приведенном определении воля является антагонистом мотивов поведения (инстинкта, привычек). В этом смысле воля противоположна желаниям (их рациональной составляющей) и характеру (как совокупности типичных для заданного индивидуума мотивов).

С физической точки зрения мотивам можно сопоставить волновую функцию, определяемую текущей конфигурацией нервных связей и импульсов. Возможно, такая разновидность мотивов, как инстинкт, отражена в самой структуре мозга (в порядке сцепления нейронов и в их составе). Глубоко укоренившиеся привычки так же, наверное, могут закрепляться в структуре. Но мотив — это ещё и образ (представление), который, по-видимому, соответствует текущим электрохимическим импульсам.

Итак, можно резюмировать, что мотивам на физическом уровне соответствует сама структура мозга и вся совокупность электрохимических импульсов, которые в нём протекают. На долю воли при этом остаются квантовые эффекты, присутствие которых мы считаем неизбежным при работе мозга. Основной вопрос этого раздела состоит в том, есть ли принципиальное отличие волевых актов человека от действий "шумящего" робота.

С точки зрения квантовой механики различия нет: как одно, так и другое являются абсолютно случайными процессами. При этом вопрос, как же все-таки реализуется исход при заданной вероятности, просто не рассматривается. Личная позиция автора [Неделько, 1994] заключается в том, что реализация случайного исхода определяется некоторой сущностью, которая может обладать индивидуальностью. И волевые акты человека объединены его личностным единством. Причем эта индивидуальность никак не связана с характером.

В качестве иллюстрации для введенного понятия воли можно привести следующий пример. Рассмотрим программный генератор псевдослучайных чисел в диапазоне $[0, 1]$. Числа, которые он дает, строго говоря, не являются случайными. Однако с точки зрения обычных статистических критериев они ничем не отличаются от случайных. Теперь мы можем рассмотреть два подобных генератора (использующие различные алгоритмы). Анализируя последовательности чисел, ими произведенные, обычными статистическими методами нельзя определить, каким генератором какое число произведено. Тем не менее, от этого они не перестают быть разными генераторами. Точно так же случайность квантовых эффектов вовсе не исключает возможность того, что эти эффекты на самом деле псевдослучайны, и эта псевдослучайность различна в разных ситуациях даже при одинаковых распределениях вероятностей (волновых функциях).

Выясним теперь, какое отношение воля может иметь к интеллекту. Поскольку воля – это «индивидуальный способ реализации случайности», она может, гипотетически, действовать «против энтропии» и, вообще говоря, «не подчиняться закону больших чисел» (смысл этой взятой в кавычки фразы уточняется в следующем разделе). Такое свойство воли может быть (если имеет место) определяющим для творчества, а именно, давать возможность находить решения, которые маловероятно обнаружить случайным поиском. Видимо, в этом аспекте искусственный интеллект будет принципиально уступать естественному.

Содержательная интерпретация вероятности.

Вероятностью называют значение вероятностной меры, которая в свою очередь определяется как некоторая функция, заданная на σ–алгебре событий и удовлетворяющая аксиомам Колмогорова.

При этом теория не дает правил для задания вероятностей в практических ситуациях. Так например, вероятность падения монеты гербом можно выбрать любым числом в интервале $(0, 1)$. Даже если в качестве эксперимента проведено 1000 подбрасываний монеты, из которых 508 реализовались гербом, это не исключает возможности для вероятности быть равной, например, 0,7. Конечно, при такой гипотезе вероятность получить подобный результат эксперимента (такое же или меньшее число гербов) очень низка, но, тем не менее, ненулевая. Хотя даже нулевая вероятность события еще не означает, что событие невозможное.

Поэтому правильное задание вероятностей делается на основе не математических, а эмпирических законов. Основной эмпирический факт в теории вероятностей можно сформулировать следующим образом: в практических задачах возможно задать вероятностную меру так, что более ожидаемыми будет правильно считать более вероятные события.

При этом есть общие правила для адекватного задания вероятностей. Например, если исходы полностью симметричны (или однородны) из физических условий (например, рассматривается симметричная монета), то они должны приниматься равновероятными. Нулевая вероятность события равносильна тому, что это событие гарантированно не произойдет. При этом важно отметить, что даже если событие гарантированно не произойдет, это не значит, что оно является невозможным. Например при случайном выборе числа из равномерного распределения на интервале $[0, 1]$ можно быть уверенным, что результат не будет равен 0,5 (как и вообще любому наперед заданному значению). Однако событие выбора 0,5 не является невозможным.

Теперь укажем, в каком смысле говорилось, что воля может нарушать закон больших чисел. Разумеется, что, являясь теоремой, ЗБЧ не может нарушаться в математическом смысле, а под «нарушением» имелось в виду невыполнение именно эмпирических правил интерпретации вероятностей.

Пусть мы достоверно узнали, что вероятность выигрыша на лотерейный билет А равна 0,8, а на билет В – только 0,2. Очевидно, правильным решением будет сделать ставку на А. Если же реализация исхода определяется человеческой волей, правильным может быть ожидание события с меньшей вероятностью.

Можно было бы считать, что воля изменяет вероятности – но это не оправдано, поскольку это изменение вероятностей не выявляется статистическими средствами. Минимально достаточным будет допущение возможности для воли нарушать вероятностные предпочтения.

Согласование экспертных высказываний.

В качестве примера построения и использования полного пространства моделей для высказываний рассмотрим разработанный ранее подход к построению решающей функции на основе несогласованных вероятностных логических высказываний экспертов [Лбов, Неделько, 1997], [Неделько, 2000].

Будем использовать экспертную информацию, заданную вероятностными логическими высказываниями вида: "**Если** (*температура воздуха в* 13^{00}) ≥ 12°C **и** (*температура воздуха в* 23^{00}) ≥ 7°C **и** (*атмосферное давление*) > 755 мм. р. ст. **или** (*температура воздуха в* 23^{00}) ≤ 4°C, **то** (*заморозок*) с вероятностью 0,4; **степень доверия** высказыванию = 0,8 ".

Смысл такого рода высказываний заключается в оценке зависимости некоторой целевой переменной Y от измеряемых переменных $X_1,\ldots, X_n$. Поэтому моделями будет множество функций в некотором пространстве.

Множество допустимых значений переменной будем обозначать так же, как саму переменную и обозначим:

$$X = \prod_{j=1}^{n} X_j\,,\ Y = \prod_{j=1}^{m} Y_j\,,\ D = X \times Y\,.$$

Пусть в D определена вероятностная мера $P_c[D]$ (квадратные скобки будем использовать, чтобы отличать меру на множестве от $P(\cdot)$ – вероятности события). Для идентификации различных вероятностных мер введем множество C, элементы которого будем называть стратегиями природы и обозначать c.

Определим функцию условной вероятности как $f_c(x) = P_c(y = 1/x)$ – вероятность принадлежности первому классу при условии известного *x*. Формально, $f_c(x) = \frac{d\,P_c^1[X]}{d\,P_c[X]}$, где мера $P_c^{\omega}[X]$ определяется как $\forall E \subseteq X$, $P_c^{\omega}(E) = P_c(E \times \{\omega\})$, а $P_c[X] = P_c^1[X] + P_c^2[X]$ – маргинальная мера, соответствующая $P_c[D]$.

Оценить $f_c(x)$ на основе эмпирической информации *v*, зная ее функцию правдоподобия, можно, определив апостериорное распределение на стратегиях C, используя формулу Байеса:

$$P[C/v] = \frac{P(v/c)P[C]}{\int\limits_C P(v/c)\,d\,P[C]},$$

где $P(v/c)$ – функция правдоподобия для набора высказываний $v = \left\{B_i \;\middle|\; i = \overline{1,N}\right\}$.

Основная идея, которая используется при восстановлении $f_c(x)$ по экспертной информации, заключается в интерпретации появлений экспертных высказываний, как случайных событий, вероятность которых зависит от того, какая *c* имеет место в действительности. Если высказывания сделаны независимо, то $P(v/c) = \prod_{i=1}^{N} P(B_i/c)$.

Для применения статистического подхода необходимо знать, как вероятность появления заданного высказывания зависит от *c*. При этом, если пользоваться формулой Байеса, то искомую зависимость $P(B_i/c)$ достаточно знать с точностью до множителя, не зависящего от *c*. Такую зависимость естественно называть функцией правдоподобия.

Философские вопросы.

Хотя основное содержание данной работы заключается в формулировании некоторого терминологического аппарата, необходимо хотя бы кратко коснуться его философской интерпретации. В качестве эталонного изложения философских концепций будем опираться на классический учебник [Russell, 1946].

Основной вопрос данного раздела: каким из рассмотренных понятий соответствуют сущности в реальности, и какова их подчиненность (какие первичны, какие вторичны).

Человек идентифицирует свое «Я» как сознание, которое выражается в восприятии окружающего мира. Поскольку независимого определения окружающего мира мы не давали, будем пока отождествлять его с совокупностью ощущений в восприятии. Часть этого мира является зависимым от сознания, а именно, человек может управлять своим телом. Подавляющая часть мира не зависит от сознания. Более того, даже зависимая часть имеет ограничения на управление: так человек может переместить свое тело, например, посредством ходьбы, но не телепортации. То, что (хотя бы частично) не зависит ни от чьего сознания, и назовем материей.

Из того, что материя по определению в своих проявлениях не зависит от сознания, еще не следует, что существование материи независимо от существования сознания. И одна из проблем в том, что определить само понятие существования безотносительно сознания достаточно сложно. Действительно, каждому понятно словосочетание «я существую», так же понятно существование того, что находится в ощущении. Но весьма нетривиально наделить смыслом утверждение «материя существует сама по себе», не проводя аналогий с собственным существованием (которые некорректны, если считать материю неодушевленной). Итак, можно выделить по крайней мере три различных понятия, идентифицируемых словом «существование»: существование «Я», существование ощущений и существование модели, то есть абстракции (например, числа), но ни одно из них не подходит для материи.

Для интерпретации изложенного в данной работе подхода, наиболее подходящей философской концепцией будет считать материю и волю первичными сущностями мироздания, которые не имеют

смысла друг без друга (или являются разными «сторонами» одной сущности, как квантово-волновой дуализм в физике). Тогда сознание будет эффектом взаимодействия воли и материи. А в общем случае взаимодействие воли и материи можно считать определением понятия существования в самом базовом смысле. Такая концепция может считаться материалистической, поскольку сознание здесь вторично. Правда, существование также вторично, при этом сознание является не формой существования, а частным случаем, иначе говоря, существование – это нечто, аналогичное сознанию, но более универсальное.

Введенные понятия имеют практическую значимость и в том, что позволяют проинтерпретировать ряд феноменов. В частности, ситуацию, когда индивид совершает так называемое «волевое усилие», можно объяснить тем, что принимается решение, имеющее низкую вероятность при заданной картине мозговых импульсов. Можно выдвигать и более спорные гипотезы, например возможность для чужой воли непосредственно вмешивается в управление мозговой активностью, что например могло бы быть альтернативным объяснением феномена гипноза. Подобная гипотеза, однако, не имеет научных обоснований и отмечена лишь как логическая возможность. Также понятие воли позволяет провести принципиальное различие между понятиями «я хочу» и «мне хочется», то есть между волей и характером, что имеет практическое социальное значение (хотя не дается ответа, как отличить собственную волю от приобретённых или врожденных мотивов, в частности, стереотипов). Свобода для воли в этом контексте означает не отсутствие необходимости, а наличие индивидуальности.

Кроме того, согласно введенному определению воля бессмертна (точнее говоря, понятие смерти к ней неприменимо). Естественно, что это не имеет ничего общего с теорией «переселения душ», поскольку допускается лишь то, что два последовательно живущих индивида «управляются одним и тем же генератором псевдослучайных чисел», но не передача какой-либо информации между ними.

Заключение

Основная идея данной работы заключается в формализации понятия смысла высказывания посредством введения пространства моделей. В частности это позволяет для анализа текстов на естественном языке выбрать математический аппарат, который был бы достаточно строгим, но более гибким и наглядным по сравнению с теориями формальных языков.

Поскольку использование языковых средств является важнейшим атрибутом интеллектуальной деятельности, естественным образом в рассмотрение оказался вовлечен широкий круг сопутствующих понятий.

Литература

[Chang, Keisler 1973] C.C. Chang, H.J. Keisler. Model Theory. / Studies in Logic and Foundations of Mathematics. Vol. 73. London. 1973. – Г. Кейслер, Ч.Ч. Чен. Теория моделей. М.: Мир. 1977. 612 с.

[Викентьев, 2004] А.А. Викентьев. Метрика и информативность на знаниях экспертов в различных моделях теорий // Искусственный интеллект т.2, НАН Украины, 2004, с.37-42.

[Лбов, Неделько, 1997] Г.С. Лбов, В.М. Неделько. Байесовский подход к решению задачи прогнозирования на основе информации экспертов и таблицы данных. // Доклады РАН. Том 357. № 1. 1997. С 29–32.

[Неделько, 2000] В.М. Неделько. Байесовская стратегия прогнозирования разнотипного временного ряда на основе выборки и экспертных высказываний. // III Международная конференции по мягким вычислениям и измерениям (SCM-2000). Сборник докладов. Июнь 2000, С. Петербург. С. 123–126.

[Шопенгауэр] А. Шопенгауэр. Избранные произведения. М. Просвещение, 1992.

[Неделько, 1994]. Проблема свободы воли в философии А. Шопенгауэра. 1994. 12 с.

[Russell, 1946] B. Russell. History of Western Philosophy. London. 1946. – Б. Рассел. История западной философии. Новосибирск. 2003. 991 с.

Информация об авторе

Виктор Михайлович Неделько *– с.н.с. лаборатории Анализа данных Института математики СО РАН, пр-т Коптюга, 4, Новосибирск, 630090, Россия, e-mail: nedelko@math.nsc.ru*