

МОДЕЛИРОВАНИЕ СЕМАНТИЧЕСКОГО ПРОСТРАНСТВА ЗНАНИЙ ИЗ РАЗНЫХ ИСТОЧНИКОВ

Abstract. (The article describes the semantic model of the domain (semantic space) for the tasks of analysis and meaningful interpretation of monitoring data of air. Proposed visualization of environmental indices in a two-dimensional. scales).

Введение. Базы знаний в экспертных системах обычно отражают лишь поверхностную составляющую знаний эксперта, не затрагивая их глубинной сути. Для извлечения более глубоких знаний предлагается использовать методы *психосемантики*, которые ориентированы на анализ структуры знаний эксперта при помощи моделирования его индивидуального опыта и выявления глубинных (неосознанных) структур знаний [1, 2]. Эти методы успешно применяются при разработке баз знаний интеллектуальных систем, обеспечивая возможность реконструировать семантическое пространство памяти для моделирования поля знаний эксперта.

Поле знаний будем называть неформальное описание понятий и взаимосвязей исследуемой предметной области, которое составляется при работе с экспертом и описывается в виде графа, диаграммы, таблицы или текста. Под моделированием семантического пространства знаний эксперта понимается переход к формализованному описанию этих знаний.

Семантическое пространство знаний может содержать как экспертные оценки, так и информацию, полученную на основе обработки результатов измерений и экспериментальных данных из разных источников.

Семантическое поле. Вероятностное представление семантического поля предложено В.В. Налимовым [3]. У Налимова воспринимаемый нами окружающий мир интерпретируется как множество текстов. Эволюция текстов связывается со спонтанным появлением фильтров, изменяющих систему восприятия. Процесс изменения отражает формула Байеса.

В.В. Налимов предположил, что все возможные интерпретации смыслов упорядочены вдоль шкалы и соотнесены с линейным континуумом Кантора – числовой осью μ . С такой позиции мир можно рассматривать как еще не проявленный семантический вакуум, который раскрывается через процесс придания смыслов его текстам. Для проявления смыслов производится вероятностное взвешивание, и разным участкам шкалы приписывается разная мера. Семантика каждого текста задается своей функцией распределения $p(\mu)$, которую можно сопоставить с окном в семантический мир. Изменение текста Налимов связывает со спонтанным возникновением в новой жизненной ситуации y фильтра $p(y/\mu)$:

$$p(\mu/y) = k p(\mu) p(y/\mu),$$

где $p(\mu/y)$ – функция распределения для семантики нового текста после события y , k – константа нормирования.

Видимая нам индивидуальность человека (Эго) задается плотностью вероятности, построенной на семантической шкале (рис.1).

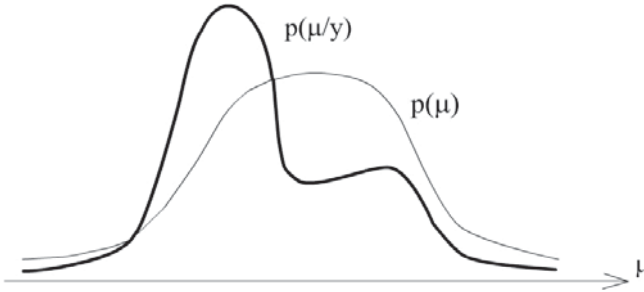


Рис. 1. Примеры семантических портретов

В процессе становления личности эта функция может существенно меняться, определяя индивидуальные особенности психики. Следовательно, на информационном уровне *индивидуальность* следует рассматривать как присущий данной конкретной личности способ интерпретации жизненных ситуаций, корректируемый в процессе индивидуальной жизни.

Функция распределения, заданная указанным способом, синтезирует в себе способ восприятия и способ суждения, математически описывая процесс перераспределения смыслов, связанный с возникновением новых сложных ситуаций. Фактически речь идет об изменениях в структуре ассоциативных связей. При этом можно воспользоваться понятием *установки*, которая в данном случае имеет динамический характер. Происходит взаимодействие исходной смысловой установки с корректирующим фильтром, в результате чего принимается то или иное решение [3, с.180].

Отметим, что одномерный случай носит в основном иллюстративный характер, поскольку реальная личность, как правило, совмещает в себе несколько ценностных критериев, которые могут вступать между собой в противоречие. Одномерная функция распределения в некоторой степени отражает представление о целостной личности, образуя пик или плато в области основного смысла жизни (служения идее, творческой активности).

В других случаях она выглядит как ряд подъемов и спадов, отображая различные группы смысловых ценностей. В общем случае будем говорить о текстах, задаваемых в многомерном семантическом пространстве.

На формальном уровне *семантическое пространство знаний* будем рассматривать как совокупность определенным образом организованных признаков (параметров), описывающих и дифференцирующих объекты (смыслы или значения) некоторой предметной области. При этом следует

указать определенное правило группировки этих признаков в более емкие категории, которые выступают в роли понятий для описания отдельных значений, оценки сходства или других отношений между ними.

В двумерном случае смысловые шкалы можно изображать на плоскости, что позволяет перейти к визуальной интерпретации некоторого множества значений или событий. Многомерное пространство знаний можно описать с помощью нескольких информативных проекций.

Методы моделирования. Для моделирования семантических пространств экспертных знаний в работах [1, 2] предлагается использовать методы многомерной статистики. В большинстве случаев используются методы факторного анализа, многомерного шкалирования или кластерного анализа. Эти методы позволяют группировать отдельные признаки описания объекта в более емкие категории – факторы. На языке поля знаний факторы являются концептами более высокого уровня абстракции. При геометрической интерпретации такого пространства значение отдельного признака отображается как точка или вектор с заданными координатами в n -мерном пространстве, координатами которого служат выделенные факторы. Таким образом, построение семантического пространства предполагает переход к описанию предметной области на более высоком уровне абстракции, то есть переход от языка с большим количеством признаков к более емкому языку концептуализации, который в данном случае выступает как метаязык по отношению к первому.

В зависимости от личного опыта и профессиональной компетентности специалиста размерность его семантического пространства и расположение в нем тех или иных понятий может существенно меняться. Эти свойства семантических пространств можно использовать для контроля процесса обучения, при тестировании экспертов или пользователей [1]. При анализе индивидуального семантического пространства выявляются те вопросы, которые не были усвоены и систематизированы. Для контроля производится сравнение семантического пространства испытуемого со структурами знаний опытных специалистов. В соответствии со степенью согласованности этих пространств, их размерности и конфигурации понятий можно определить уровень знаний и подготовки испытуемого.

При использовании математических методов преобразования сложных структур данных в более простую форму различные методы приводят к разным структурам. В частности, кластерный анализ порождает древовидные структуры, факторный анализ и методы многомерного шкалирования – пространственную интерпретацию множества точек.

Методы шкалирования позволяют выявлять структуры знаний косвенным путем, при получении ответов на довольно простые вопросы о близости или различии между понятиями X и Y. Многие эксперименты подтверждают такую закономерность: с повышением профессионального уровня специалиста размерность его семантического пространства

уменьшается. Этот вывод вполне согласуется с известным в когнитивной психологии положением, что процесс познания ведет к обобщению.

Процесс моделирования семантического пространства на основе методов многомерной статистики включает три этапа [2]:

1. Выбор и применение подходящего метода оценки семантического сходства. Этот шаг включает эксперимент, в процессе которого специалисту предлагается оценить общность предъявляемых признаков с использованием некоторой шкалы.

2. Построение семантического пространства на основе статистического анализа полученной матрицы сходства. При этом происходит уменьшение количества исследуемых понятий (параметров) за счет перехода к более общим координатам (факторам или индексам).

3. Поиск смысловых (языковых) эквивалентов для выделенных структур: идентификация или интерпретация полученных факторов, кластеров, групп объектов или координатных осей.

Представляя форму обобщения информации, значение может также выступать как оператор классификации, упорядочивающий объекты или события исследуемой области. Результатом семантического анализа будет выявление семантических связей между значениями, которые можно представить в свернутом виде. Формой фиксации семантических связей могут быть семантические поля, тезаурусы или онтологии.

Динамика семантических структур. Семантические пространства отражают семантическую организацию различных уровней репрезентации знаний (уровней категоризации). Эти уровни соединены друг с другом и определяются степенью общности тех признаков, на основе которых производится выделение понятий. Наиболее глубокий уровень обобщения основан на механизмах синестезии, которые связывают ощущения разных модальностей (например, звуки музыки и цветовые образы). Это подтверждается данными о совпадении вербальных и образных семантических структур на уровне глубокой семантики.

Процесс порождения актуальных значений можно представить как последовательную развертку содержания воспринятого субъектом материала в ее движении от эмоционально-образных форм к уровням предметной расчлененности. При моделировании с помощью семантических пространств этот процесс представляется как последовательная развертка небольшого числа наиболее общих факторов в совокупности более конкретных характеристик. Соответственно, размерность семантического пространства, построенного для конкретной предметной области, содержание выявленных факторов и их взаимосвязей прежде всего зависит от субъективного знания данной предметной области.

Изучение какой-либо области способствует переходу от более простых семантических структур к дифференцированным предметно-категориальным пространствам. В то же время, в ряде работ показано, что эмоции могут влиять на семантические структуры, а аффекты автоматически приводят к

уменьшению размерности (за счет перехода на глубинные уровни). Обычно такой переход имеет адаптивный смысл, так как позволяет установить более широкие ассоциативные связи между разными областями. Обладая сходной эмоциональной окраской, объекты оказываются близкими на глубинных уровнях и образуют синкретические обобщения, присущие мифологическому сознанию, феноменологии сна или творческого процесса.

Следовательно, семантическое пространство можно представить как подвижную функциональную систему, которая меняет свою размерность в зависимости от той задачи, которую решает субъект в данных момент, обращаясь к категориальным структурам разных уровней.

Семантическое пространство риска. Во многих прикладных задачах семантическое значение сложной ситуации может быть связано с наличием или отсутствием опасности. Поэтому в качестве критерия для уточнения степени опасности рассмотрим вероятностное определение риска [4, 5].

Каждая вероятностная модель описывает некоторое множество наблюдений X , на котором определено разбиение $E(X)$, которое образует множество случайных событий (ситуаций). На этом множестве задается неотрицательная функция P , которая удовлетворяет известным свойствам вероятностной меры и определяет вероятность случайных событий из $E(X)$. Множество наблюдений X , его разбиение $E(X)$ и вероятностная мера P , заданная на этом разбиении, образуют пространство риска (X, E, P) .

Само понятие риска обусловлено вероятностным представлением о воздействии факторов среды на состояние исследуемой системы. Величину риска нельзя непосредственно измерить, а можно лишь с некоторой степенью надежности оценить, используя количественные характеристики факторов риска и данные об их воздействии. Факторами риска могут быть факторы любой природы (физические, генетические, социальные и др.), имеющие отношение к нарушениям гомеостаза системы. Количественная оценка достоверности эмпирически установленных связей производится методами математической статистики, которые позволяют учесть неопределенности, возникающие как при измерении самих воздействий, так и при оценке связи между воздействием и эффектом.

Если в качестве событий A_1, A_2, A_3, \dots рассматриваются различные последствия воздействия некоторого фактора, то значения вероятностей для разных исходов служат характеристиками риска, обусловленного данным фактором. В любой конкретной ситуации вероятности отдельных исходов будут удовлетворять известным свойствам вероятностной меры, т.е. вероятность любого события неотрицательна, вероятность всех исходов составляет 1, вероятность суммы непересекающихся событий равна сумме вероятностей этих событий. Например, если в качестве возможных исходов принять степени тяжести конкретного заболевания, то для любой степени тяжести величина риска попадет в интервал от 0 до 1, а риск суммы двух исходов равен сумме двух рисков. Более детально основные этапы и методы моделирования пространства риска представлены в работах [5 - 6].

Чтобы определить риск для здоровья, необходимо исследовать множество разных факторов, среди которых важную роль играет наследственность (генетический риск), социально-экономические условия (социальный риск), образ жизни, медицинское обслуживание и качество окружающей среды. С каждым из этих факторов можно сопоставить более конкретные виды рисков, т.е. от общих представлений перейти к более конкретным показателям, которые можно оценить с помощью экспертов или непосредственных измерений. Так, генетический риск охватывает риски для разных групп заболеваний (онкологические заболевания, заболевания сердца и сосудов, психические заболевания и др.). Риск, связанный с образом жизни, может возрастать за счет профессионального риска, неправильного питания, вредных привычек, увлечения экстремальными видами спорта.

Экологический риск – это риск, который обусловлен загрязнением атмосферы, питьевой воды и продуктов питания. В свою очередь, каждый из этих факторов можно конкретизировать, указывая количественные оценки вредных примесей (концентрации в атмосферном воздухе, воде и т.д.).

В теоретических исследованиях семантическое пространство риска строится «сверху вниз», т.е. выбираются наиболее общие факторы риска, которые затем разворачиваются в отдельные показатели и конкретизируются в виде количественных оценок.

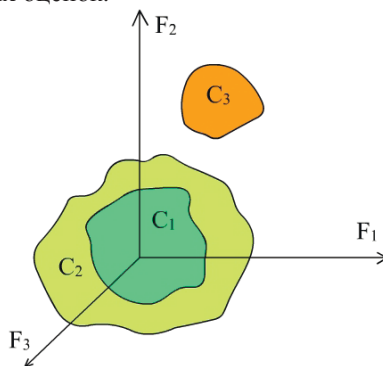


Рис. 2. Пример пространства риска для здоровья

При анализе экспериментальных данных моделирование происходит в обратном порядке, от конкретных измерений или оценок к обобщенным показателям (индексам). Полученное в результате систематизации натуральных измерений и экспертных оценок множество количественных показателей может быть свернуто в небольшое число индексов (факторов риска), которые образуют координатные оси семантического пространства.

На рис. 2 показано пространство риска, построенное по трем группам показателей, где F1 – генетический риск, F2 – риск, обусловленный образом жизни, F3 – экологический риск. В этом примере факторы F1 и F2

определяют риск, обусловленный особенностями конкретных индивидов (групп риска). Фактор F3 представляет социально-экологический фактор риска, связанный с техногенными нагрузками на отдельный регион. Область C1 характеризуется достаточно низким уровнем риска, C2 – область повышенного риска, C3 – область высокого риска.

В тоже время, при оценке риска и принятии решения необходимо учитывать субъективный фактор, связанный с личным опытом эксперта. Каждое новое событие требует принятия решения с учетом накопленного опыта, но и само решение (точнее, его последствия) влияет на структуру сформированного ранее пространства знаний, т.е. каждый шаг приводит к очередной коррекции этих знаний и влияет на принятие решений в будущем.

Повышенная индивидуальная склонность к риску может формироваться в результате приобретения жизненного опыта в тех ситуациях, когда рискованное поведение было оправдано. Отрицательный жизненный опыт в подобных ситуациях приводит к стратегии минимизации риска. Если два индивида с различным жизненным опытом попадут в одну и ту же ситуацию, где выигрыш не зависит от риска, то их поведение все равно уже обусловлено опытом адаптации к разным классам ситуаций. Механизм влияния установки на стратегии поведения был формализован в моделях динамического уровня адаптации [7].

На рис. 3 показаны субъективные представления об успехе для двух индивидов с разным жизненным опытом. В результате в нейтральной ситуации S индивид N1 выберет неоправданно рискованное решение, а индивид N2 будет проявлять чрезмерную осторожность.

Многие исследования подтверждают, что склонность к риску не является постоянной характеристикой личности, поскольку один и тот же человек может демонстрировать разные стратегии поведения в соответствии с имеющимся опытом для разных классов ситуаций.

Выводы. Рассматриваются методы моделирования семантического пространства знаний, предложенные в психологических исследованиях. Эти методы можно адаптировать для моделирования пространства экспертных знаний при разработке интеллектуальных систем принятия решений.

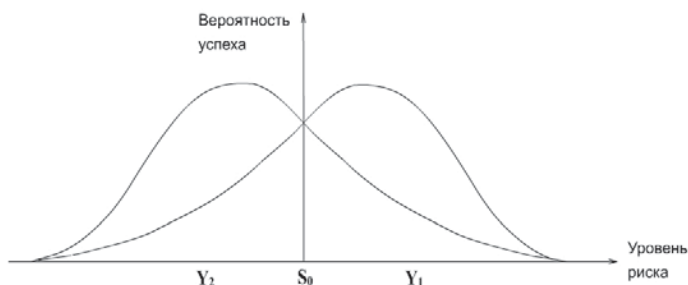


Рис. 3. Влияние субъективного опыта на выбор решения.

Предложен общий подход к моделированию пространства знаний эксперта «сверху вниз», где на первом этапе рассматриваются наиболее общие факторы (понятия, категории), а последующие этапы направлены на разворачивание и конкретизацию этих факторов до уровня отдельных количественных показателей или вероятностных оценок. Рассматривается пример построения семантического пространства риска для здоровья, которое включает как внешние факторы влияния (негативные аспекты окружающей среды), так и индивидуальные особенности (показатели уязвимости) для отдельных групп населения.

Предложенный подход может использоваться при организации базы экспертных знаний и оценивания рисков в тех случаях, когда необходимо анализировать большие объемы статистической информации и принимать решения с учетом субъективного опыта, приобретенного экспертами.

1. *Гаврилова Т.А., Хорошевский В.Ф.* Базы знаний интеллектуальных систем. – СПб.: Питер, 2001. – 384 с.
2. *Петренко В.Ф.* Психосемантика сознания. М.: Изд-во МГУ, 1988. – 230 с.
3. *Налимов В.В.* Спонтанность сознания: Вероятностная теория смыслов и смысловая архитектура личности. - М.: «Прометей», 1989. – 288с.
4. *Большаков А. М., Крутько В. Н., Пуццлло Е. В.* Оценка и управление рисками влияния окружающей среды на здоровье населения. – М.: Эдиториал УРСС, 1999. – 256 с.
5. *Каменева И.П.* Вероятностные модели репрезентации знаний в интеллектуальных системах принятия решений // Искусственный интеллект. - 2005. - № 3. С. 399-409.
6. *Каменева І.П.* Просторово-семантичні моделі репрезентації знань в геоекологічних дослідженнях // Геоінформатика. - 2005. - № 4. С. 64-69.
7. *Шмелев А. Г.* Психодиагностика личностных черт. СПб: Речь, 2002. – 480 с.

Поступила 2.10.2013р.

УДК 504.064:303.732:519.7

О.О.Попов, м. Київ

КОНЦЕПТУАЛЬНО-МЕТОДОЛОГІЧНІ АСПЕКТИ МОДЕЛЮВАННЯ ВПЛИВУ ОБ'ЄКТІВ АТОМНОЇ ЕНЕРГЕТИКИ НА ДОВКІЛЛЯ

Abstract. The author examined conceptual methodological aspects of modeling large dynamic system „nuclear station – the environment” on the basis of a systems analysis. The questions of construction (selection) of mathematical models of the spread of pollutants in the subsystems of the environment and the accuracy of the calculations are considered. A list of typical models used to describe the given dynamical system is presented.