УДК 519. 816

Н.К. Тимофеева

СИТУАЦИЯ НЕОПРЕДЕЛЕННОСТИ, ВОЗНИКАЮЩАЯ В ЗАДАЧАХ СЕМАНТИКИ, И СПОСОБЫ ЕЕ РЕШЕНИЯ

Ключевые слова: задачи семантики, комбинаторная оптимизация, кластеризация, самонастраивающиеся алгоритмы, линейная свертка, меры сходства, подклассы разрешимых задач.

Введение

Одной из ключевых проблем в теории принятия решений является нахождение оптимального результата в условиях неопределенности. Эта ситуация в том или ином виде возникает в процессе решения значительной части прикладных задач, в том числе и задач семантики, и является общим случаем, а принятие оптимальных решений без ее учета есть частным случаем. Ниже рассматриваются различные виды неопределенности и способы их решения, которые возникают при решении задач комбинаторной оптимизации. Поскольку задачи семантики сводятся к задачам данного класса, то описанные виды неопределенности и способы их решения справедливы и для семантики.

Постановка проблемы и цель исследования

Для нахождения в задачах семантики в условиях неопределенности оптимального результата, который удовлетворяет цели исследования, необходимо провести анализ ситуации и установить ее причину. Это дает возможность определить подход к ее устранению путем использования известных подходов или разработки новых.

Анализ последних исследований и публикаций

В языкознании (философии) имеют место понятия «определенный» и «неопределенный». В семантике оперируют термином «предмет» (объект), который по условию является определенным. Считают, что неопределенность — это воспроизведение предметов и явлений действительности, при котором актуализируются только их общие признаки [1].

В данной работе затрагивается присущая теории принятия оптимальных решений проблема неопределенности, которая возникает при решении оптимальных задач различной природы, в том числе и семантики. Это понятие используется на интуитивном уровне. Для оценки степени такого состояния объекта вводится мера неопределенности [2]. Величина, характеризующая количество неопределенности в теории информации, называется энтропия, точнее, информационная энтропия [2].

В теории принятия решений, в основном, исследуют неопределенность, связанную с неполной входной, текущей и нечеткой информацией [2–5]. Но эта ситуация, в частности в комбинаторной оптимизации, к которой сводятся задачи семантики, проявляется благодаря особой структуре множества комбинаторных конфигураций, которые являются аргументом целевой функции, и способу ее моделирования [6]. В данном случае достаточно сложно смоделировать такую целевую функцию, чтобы глобальное решение совпадало с реальной целью исследования. Задачи семантики относятся к задачам распознавания и для установления сути определенного объекта требуют нахождения его эталона в базе данных. Для сравнения заданного объекта с эталоном вводятся меры сходства, которые являются субъективными оценками. Возникает ситуация неопределенности вследствие выбора среди нескольких оптимальных вариантов решений, полученных с использованием различных мер сходства (или разработанных целевых функций), такого, который удовлетворял бы цели исследования. Далее рассмотрим некоторые виды неопределенности, имеющие место в комбинаторной оптимизации.

Классификация неопределенности в задачах семантики

Задачи семантики требуют установления сути определенного объекта. Для этого необходимо найти соответствующий эталон в базе данных и сравнить его с заданным объектом. Путем распознавания устанавливается его суть. К задачам семантики отнесем распознавание образов, речи, детского, женского, мужского голосов, многодикторской речи, задачу клинической диагностики, сравнение текстов в целях установления плагиата, криптографию, дешифрование забытых письменностей, перевод текстов с одного языка на другой. Если при распознавании речи устанавливается сходство входного сигнала со смысловым значением слов (предложения), то это является задачей семантики.

Уточним такие понятия, как критерий и целевая функция [7].

Критерий — это признаки или свойства, которые характеризуют определенный объект и являются входными данными.

Целевая функция — выражение, которое формулируется на основе заданных критериев с учетом специфики задачи и по которому вычисляется и оценивается результат решения задачи.

Как правило, целевую функцию отождествляют с критериями. Но для одних и тех же критериев ее можно смоделировать по-разному. Если по различным смоделированным выражениям находим одно и то же решение, то можно сказать, что оно совпадает с целью исследования. В противном случае имеем ситуацию неопределенности вследствие неоднозначного результата. Возникает проблема выбора одного из нескольких оптимальных полученных по разным целевым функциям вариантов решения задачи, который совпадает с целью исследования.

В задачах семантики неопределенность связана с [6]:

- неоднозначностью неудовлетворяющего цели исследования результата, полученного по смоделированной целевой функции или выбранного мерой сходства в случае нечеткой входной информации;
- выбором способа оценки точности работы определенного алгоритма;
- особой структурой множества комбинаторных конфигураций, которые являются аргументом целевой функции;
- неполной входной, текущей или нечеткой информацией;
- нечетко разработанными правилами обработки и оценки информации;
- неоднозначностью при выборе оптимального решения по нескольким критериям в многокритериальной оптимизации.

Сформулируем такое определение.

Определение. Под неопределенностью в теории комбинаторной оптимизации понимаем ситуацию, при которой вследствие нечеткой или неполной входной и текущей информации по выбранной мере сходства, смоделированным целевым функциям, разработанным правилам обработки и оценки информации из-за особой структуры аргумента целевой функции полученный оптимальный результат не совпадает с целью исследования.

Способы решения прикладных задач в условиях неопределенности

В зависимости от вида неопределенности решение прикладных задач в условиях неопределенности выполняется различными способами. С этой целью разрабатываются самонастраивающиеся алгоритмы, в которых входные данные содержат формальные параметры, в процессе решения задачи генерируются вспомогательные критерии, с помощью которых формируется вспомогательная текущая информация с прогнозированием будущих результатов. Также используются подклассы разрешимых задач и проводится структуризация эталонной библиотеки для сведения неразрешимых задач к разрешимым. Рассмотрим подробнее эти подходы.

Вычислительная сложность задач семантики и сведение неразрешимых задач этого класса к разрешимым

Неопределенность вследствие неоднозначного, не удовлетворяющего цели исследования результата, полученного по смоделированной целевой функции или выбранной мере сходства в случае нечеткой исходной информации, появляется при распознавании определенных объектов. Введенные меры сходства в задачах семантики являются субъективной оценкой. Для различных мер оптимальное решение задачи может отличаться. Соответственно, при выборе результата возникает ситуация неопределенности. Но по некоторым мерам сходства можно найти и глобальное решение. Такие задачи выделяются в подклассы разрешимых.

Таким образом при распознавании, кроме количества операций, затраченных на нахождение глобального решения, необходимо учитывать и меры сходства, которые здесь играют основную роль и от выбора которых в значительной степени зависит само решение. Вычислительная сложность в этих задачах с учетом степени сходства оценивается по шкале «да» или «нет», где «да» означает, что выбранный способ оценки сходства позволяет найти глобальное решение, а «нет» — данный способ оценки сходства не дает ни одного решения.

Итак, по способу определения сложности решения разделим задачи на следующие типы: 1) входные данные заданы в количественном измерении, а оценка вычислительной сложности проводится по количеству операций для нахождения глобального решения; 2) входные данные заданы в качественном измерении, а оценка вычислительной сложности проводится как по числу операций, выполненных для нахождения глобального решения, так и по способу оценки качества решения (введенных мер сходства).

Решение задач распознавания требует определенной меры сходства. Обозначим $g^{+}(x, y)=1$ меру сходства, с помощью которой получаем глобальное решение. Если $g^{-}(x, y)=0,$ то по выбранной мере сходства не находим ни одного решения. Если $g_j(x, y)\in\{\chi,..., \delta\},$ то выбранная мера сходства позволяет найти решение, которое удовлетворяет цели исследования; x — объект, который необходимо распознать, y — эталонный элемент, $\chi<1, \delta>0$ — граничные оценки, для которых существует допустимое решение. Таким образом, если для определенной задачи $g^{+}(x, y)=1,$ то она является разрешимой по признаку сходства.

Приведем такой пример. В задачах распознавания проводится распознавание объектов различной природы (речевые сигналы, электрокардиограммы, электроэнцефалограммы), которые являются входными данными. Сложность решения задачи распознавания речи заключается в том, что произнесенные много раз одним и тем же или разными дикторами сигналы, соответствующие одному и тому же слову, отличаются между собой. Таких вариантов может быть бесконечное количество, потому что элементы звука (фонемы) формируются комбинацией элементов речевого тракта и образуют комбинаторную конфигурацию — размещение с повторениями. Для их распознавания вводятся меры сходства.

Представим размещение с повторениями упорядоченным множеством $\mu^{\mathrm{k}} = (\mu_1^k, ..., \mu_{\eta^k}^k) \subset W$, где $\mu_s^k \in \mu^k$ — элементы фонемы (элементы речевого тракта), $\eta \in \{1, ..., n\}$ — количество элементов в μ^k. Верхний индекс k ($k \in \{1, ..., q\}$) в μ^k обозначает порядковый номер μ^k в W, $W = \{w^k\}_1^q$ — множество разбиений с повторениями (комбинаторных конфигураций), q — количество μ^k в W.

По структуре входных данных среди речевых сигналов можно выделить фонемы, которые не пересекаются, т.е. $\mu^{\mathrm{k}} \bigcap \mu^i = \varnothing$, $k \neq i$, и не содержат одинаковых элементов. Для них мера сходства $g^+(\mu^k, \mu^i) = 1$. Задачу распознавания таких фонем назовем разрешимой как по признаку сходства, так и по структуре входного сигнала. Будем считать, что они составляют подклассы разрешимых задач распознавания речи. Если $\mu^{\mathrm{k}} \bigcap \mu^i \neq \varnothing$, то μ^k и μ^i содержат одинаковые элементы (их может быть больше одного). Для них мера сходства $g_j(\mu^k, \mu^i) \in \{\chi, ..., \delta\}$. Задача поиска прямым перебором соответствующего определенному слову в словаре речевого сигнала является NP-полной. Но она становится полиномиально разрешимой, если по определенным признакам проведено структурирование библиотеки эталонных сигналов.

Грамматические правила в синтезе речевых сигналов выступают как меры сходства. Поскольку они разработаны основательно, а человек воспринимает широкий диапазон речи, то при отсутствии условия воспроизведения индивидуальности голоса эта задача является разрешимой. Женский, мужской или детский голос зависит от частоты основного тона, амплитуды сигнала, поэтому эта задача также может быть решена в реальном времени, т.е. является разрешимой по признаку сходства. Воспроизведение индивидуального голоса зависит от гораздо более сложных факторов: от речевого тракта индивидуума, от его эмоционального состояния, от особенности его психики и т.д. Поскольку эти параметры смоделировать достаточно сложно, воспроизведение индивидуальной речи относится к классу NP-полных задач по признаку сходства.

Использование линейной свертки при решении задач семантики

Задачи семантики имеют переборный характер, поэтому они сводятся к задачам комбинаторной оптимизации. В задачах кластеризации и клинической диагностики, которые относятся к задачам искусственного интеллекта, имеет место неопределенность, связанная со структурой аргумента целевой функции.

Приведем общую математическую постановку задачи комбинаторной оптимизации и рассмотрим некоторые виды неопределенности, связанные с характерными особенностями этих задач [8]. Задачи комбинаторной оптимизации, как правило, задаются на одном или нескольких множествах, например $A = \{a_1, ..., a_n\}$ и $B = \{b_1, ..., b_{\tilde{n}}\}$, элементы которых имеют любую природу. Назовем эти множества базовыми. Имеет место два типа задач. В задачах первого типа каждое из множеств A и B можно представить в виде графа, вершинами которого являются элементы заданного множества, а каждому ребру поставлено в соответствие число $c_{sl} \in R$, которое называют весом ребра (R — множество действительных чисел, $s \in \{1, ..., n\}$, $l \in \{1, ..., \tilde{n}\}$, n — количество элементов множества A, $\tilde{n}$ — количество элементов множества B). Положим, что $n = \tilde{n}$. Числовое значение связей (весов) между элементами A и B назовем входными данными и зададим их матрицами (одна из них — комбинаторная). В задачах второго типа между элементами заданного множества связей не существует, а весами выступают числа $c_{sl} \in R$, которым соответствует некоторое свойство этих элементов.

В обоих типах задач из элементов одного из заданных множеств, например $a_s \in A$, образуется комбинаторное множество W — совокупность комбинаторных конфигураций определенного типа (перестановки, выборки различных типов, разбиения и т.д.). На элементах $w \in W$ комбинаторного множества W вводится целевая функция $F(w)$. Необходимо найти элемент w^* множества W, для которого $F(w^*)$ принимает экстремальное значение (глобальные минимум или максимум) при выполнении заданных ограничений, т.е. $F(w^*) = \text{glob} \underset{w \in W^0 \subset W}{\text{extr}} F(w)$, где $\text{extr} = \{\min, \max\}$. Элемент $w^* \in W_0 \subseteq W$ является глобальным оптимальным решением задачи комбинаторной оптимизации, W_0 — заданные ограничения.

Для задания целевой функции в явном виде и сведения ее к одному выражению для различных классов задач комбинаторной оптимизации входные данные смоделируем конечными последовательностями.

Представим элементы h наддиагоналей симметричной комбинаторной матрицы $Q(w^k)$ комбинаторной функцией $\beta(f(j), w^k)\big|_1^m = (\beta_1(f(1), w^k), ..., \beta_m(f(m), w^k))$, а элементы h наддиагоналей симметричной матрицы C — функцией натурального аргумента $\varphi(j)\big|_1^m = (\varphi(1), ..., \varphi(m))$, где $m = \dfrac{n(n-1)}{2}$ — количество элементов h наддиагоналей матриц C и $Q(w^k)$, $h = \overline{1, n-1}$, k — порядковый номер в упорядоченном множестве W. Если матрицы $Q(w^k)$ и C несимметричные, то $\beta(f(j), w^k)\big|_1^m$ и $\varphi(j)\big|_1^m$ содержат все их элементы, а $m = n^2$ (или $m = n\tilde{n}$). Суммарное произведение значений комбинаторной и числовой функций запишем как

$$F_1(w^k) = \sum_{j=1}^{m} \beta_j(f(j), w^k)\varphi(j). \qquad (1)$$

Среднее значение весов между элементами базового множества вычислим по выражению

$$F_2(w^k) = \frac{\left(\sum_{j=1}^{m} \beta_j(f(j), w^k)\varphi(j)\right)}{m}. \qquad (2)$$

Для подмножеств изоморфных комбинаторных конфигураций (комбинаторные конфигурации, в которых количество элементов (блоков) одинаковое, называются изоморфными) среднее значение принимает такой вид:

$$F_3(w^k) = \sum_{p=1}^{\eta^k} \left(\sum_{j=1}^{m} \beta_j (f(j), w_p^k)\, \varphi(j) \right) / J_p^k, \tag{3}$$

где $J_p^k = \dfrac{\xi_p^k\,!}{(\xi_p^k - 2)!2!}$, $\xi_p^k > 1$, — количество единиц в комбинаторной функции для p-го подмножества (блока) $w_p^k \subset w^k$, ξ_p^k — количество элементов в w_p^k, η^k — количество блоков (элементов) в w^k.

Выражения $F_1(w^k)$, $F_2(w^k)$, $F_3(w^k)$ являются целевыми функциями.

Рассмотрим структуру множества W. Подмножество $W_\eta \subset W$ назовем подмножеством изоморфных комбинаторных конфигураций, если его элементы — изоморфные комбинаторные конфигурации. Множество W состоит из подмножеств изоморфных комбинаторных конфигураций W_η. На подмножестве W_η целевая функция изменяется так же, как на множестве перестановок. Упорядочим подмножества $W_\eta \subset W$ (кроме перестановок), начиная с $\eta = 1$ и заканчивая $\eta = n$.

Теорема 1. Если в задачах комбинаторной оптимизации множество W состоит из подмножеств W_η, а оптимизация проводится по выражению (1), то целевая функция на заданном выше упорядочении W_η в W — дискретная кусочно-монотонная функция (соответственно неубывающая или невозрастающая).

Следствие. На множестве перестановок и подмножестве изоморфных комбинаторных конфигураций при использовании целевой функции (1) ситуация неопределенности сводится к минимуму.

Доказательство теоремы 1 приведено в [8].

Как следует из теоремы 1, ситуация неопределенности, связанная со структурой комбинаторных конфигураций, возникает вследствие того, что их множество состоит из подмножеств изоморфных комбинаторных конфигураций и на определенном их упорядочении закономерность изменения значений смоделированной целевой функции одинакова независимо от входных данных, а результат решения задачи неоднозначный.

Для выхода из ситуации неопределенности в задачах комбинаторной оптимизации, в частности кластеризации и клинической диагностики, связанной со структурой аргумента целевой функции, необходимо вводить несколько целевых функций или проводить оптимизацию по нескольким критериям, которые сводятся к взвешенному критерию (линейной свертке). Нахождение оптимального решения проводится самонастраивающимся алгоритмом с учетом постоянных и переменных критериев, которые вводятся в процессе решения задачи [9]. Иными словами, в процессе работы алгоритма генерируется дополнительная текущая информация (критерии качества), которая влияет на прогнозирование будущих результатов.

Для определения количества как прямых, так и косвенных связей между объектами, которые распознаются, вводятся матрицы, содержащие постоянную и сгенерированную информацию. Если использование введенных одноразовых критериев приводит к равнозначным вариантам решения задачи, то для устранения ситуации неопределенности вводятся переменные критерии, которые учитываются в процессе решения задачи многократно. Для каждого частичного варианта разбиения w в итерационном режиме формируется матрица, по которой вычисляется значение сформулированной целевой функции. В качестве критерия выбора при определении пары кандидатов на включение в некоторое подмножество разбиения w запишем взвешенную целевую функцию (линейную свертку) $F(w)=\sum_{t=1}^{r}\gamma_t\Phi^{(t)}(w)$, где $\gamma_t \geq 0$, а $\sum_{l=1}^{p}\gamma_t=1$ — весовые коэффициенты, r — количество целевых функций $\Phi^{(t)}(w)$, которые моделируются на основе переменных частных критериев. Выбором значения γ_t при нахождении оптимального решения изменяется степень вклада оговоренных критериев.

Для эффективного использования линейной свертки при решении многокритериальных задач семантики по каждому критерию рассматриваем частичное решение, для которого вычисляется частичная целевая функция. Если с использованием очередного критерия возникает ситуация неопределенности, вводятся дополнительные переменные критерии. Они используются как один, так и много раз в итерационном режиме.

Использование подклассов разрешимых задач для решения ситуации неопределенности при оценке точности работы алгоритма

Если оценка результата получена в числовом измерении и известен глобальный оптимум, то точность вычисляется по выражению $\left(1-\frac{F_{\min}}{F(w)}\right)100\%$ (или $\left(1-\frac{F(w)}{F_{\max}}\right)100\%$ — в случае максимизации), где $F_{\min}$ — глобальный минимум решения задачи, $F_{\max}$ — глобальный максимум решения задачи, $F(w)$ — полученное оптимальное решение задачи по определенному алгоритму. Эксперимент показывает, что чем больше размерность задачи, тем меньше погрешность (в процентах) полученного результата по отношению к глобальному оптимуму. Но задачи комбинаторной оптимизации переборные, и на больших размерностях определение глобального оптимального решения полным перебором практически невозможно. Поэтому при оценке существующими подходами точности результата, полученного определенным алгоритмом, возникает ситуация неопределенности.

Для решения описанной выше неопределенности используем разрешимый случай, который задан двумя множествами перестановок, поданных системами (y) и (x), на которых введена целевая функция $\sum yx$ [8]. Для этих систем определены перестановки, для которых $\sum yx$ принимает наибольшее или наименьшее значение. Если элементы перестановки из системы (y) упорядочены от большего к меньшему, а (x) — от меньшего к большему, то значение $\sum yx$ является глобальным минимумом. Если элементы обеих этих перестановок упорядочены от меньшего к большему, то значение $\sum yx$ является глобальным максимумом.

Теорема 2. Значение целевой функции для задач комбинаторной оптимизации, аргументом которой является перестановка, находится в пределах $\max\limits_{w^*\in(x)} F(w^*) \geq$

$\geq F(w) \geq \min\limits_{w^{**} \in (x)} F(w^{**}),\ \ w^{*},\ w^{**} \in (x),\ \ w \in W,$ — решение (комбинаторная конфигурация) определенной задачи. В случае минимизации значение целевой функции находится в пределах $\min\limits_{w^{*} \in (x)} F(w^{*}) \leq F(w) < F^{*}$. В случае максимизации — в пределах

$$F^{*} < F(w) \leq \max\limits_{w^{**} \in (x)} F(w^{**}), \quad \text{где} \quad F^{*} = \min\limits_{w^{*} \in (x)} F(w^{*}) + \frac{\max\limits_{w^{*} \in (x)} F(w^{*}) - \min\limits_{w^{**} \in (x)} F(w^{**})}{\upsilon},$$

υ — коэффициент уменьшения области поиска оптимального решения, который уточняется в процессе работы алгоритма.

Доказательство теоремы 2 приведено в [8].

С использованием подклассов разрешимых задач находится множество значений ряда индивидуальных задач. По выражению $\left(1 - \frac{F_{\min}}{F(w)}\right) 100\%$ (или $\left(1 - \frac{F(w)}{F_{\max}}\right) 100\%$ — при максимизации) оценивается точность алгоритма (метода).

Для других типов комбинаторных конфигураций теорема 2 справедлива на подмножествах изоморфных комбинаторных конфигураций.

Использование самонастраивающихся алгоритмов для выхода из ситуации неопределенности, связанной с неполной входной и текущей информацией

Для различных прикладных задач ситуация неопределенности, связанная с неполной входной и текущей информацией, решается по-разному. В одних случаях проводится анализ поведения системы за определенный промежуток времени с последующим установлением определенной закономерности, которая учитывается при прогнозировании будущих результатов на текущем отрезке времени [2, 4]. В других случаях используют экспертные системы. Для выхода из ситуации неопределенности используем самонастраивающиеся алгоритмы [10].

Живая природа представляет собой самоприспосабливающуюся (адаптивную) систему. По аналогии с живой природой в теории управления разработаны адаптивные системы автоматического управления [11, 12]. В этих системах автоматически меняется алгоритм управления в целях сохранения показателей качества при произвольном смещении характеристик управляемого объекта. Самонастраивающаяся система управления может учитывать не только информацию, но и прошлый опыт. В этом случае добавляется блок оперативной памяти, в котором накапливаются сведения об управляемом технологическом процессе. Коррекция программы проводится на основании обобщения опыта работы машины-автомата. По характеру изменений в управляющем устройстве адаптивные системы разделяют на две большие группы:

- самонастраивающиеся, в которых меняются только значения параметров регулятора.
- самоорганизующиеся, в которых меняется структура самого регулятора.

Для разрешения различных видов ситуации неопределенности в комбинаторной оптимизации используем самонастраивающиеся алгоритмы. В процессе своей работы они автоматически изменяют те или иные параметры или генерируют дополнительную информацию в целях нахождения оптимального решения в определенных условиях.

Прикладные задачи комбинаторной оптимизации (или семантики), как правило, сложные по своей природе и разбиваются на подзадачи, для решения которых разрабатывают независимые алгоритмы. Основная задача решается последовательной работой этих алгоритмов или они работают как встроенные процедуры в итерационном режиме. В процессе их работы при переходе от решения одной задачи к другой в момент передачи информации, которая является результатом решения предыдущей, на входе другого алгоритма появляются новые, неопределенные параметры, которые необходимы для решения следующей задачи и которые невозможно задать во входных данных по условию. Возникает проблема нахождения параметров в условиях неопределенности.

Таким образом, для полной автоматизации процесса нахождения оптимального результата разрабатываются самонастраивающиеся алгоритмы генерации параметров, которые необходимо задавать как входные данные для решения очередной задачи и которые невозможно задать в начале вычислительного процесса [9]. Это позволяет в процессе решения определенной задачи с учетом предыдущих результатов генерировать дополнительную информацию с прогнозированием будущих результатов. В этом случае при подготовке исходных данных вводятся формальные параметры. Действительные параметры, по которым находится оптимальное решение, генерируются автоматически самонастраивающейся программой-генератором по разработанным правилам. Иными словами, для принятия оптимального решения в условиях неопределенности в них реализованы элементы прогнозирования и возвращения к переоценке предыдущего результата в автоматическом режиме, что характерно для искусственного интеллекта.

Неопределенность, связанная с нечетко разработанными правилами обработки и оценки информации

Рассмотрим данную проблему на примере клинической диагностики. Эта задача относится к задачам семантики.

Постановка диагноза проводится с использованием определенных правил обработки и оптимальной оценки информации, которыми располагает врач. Выделим следующие:

• правила, которые четко сформулированы и описаны в книгах и учебниках. Назовем их правилами обучения;

• правила выделения характерных признаков определенного заболевания и установления корректного диагноза, которые формируются в процессе практической деятельности врача (правила самообучения);

• правила выделения характерных признаков определенного заболевания без лабораторных данных и компьютерной диагностики на уровне интуиции благодаря особому мышлению врача. Этими свойствами обладает ограниченная категория людей (правила интуиции).

Правила принятия оптимального решения, касающиеся обучения и самообучения, достаточно основательны и формализуются путем введения целевой функции (одной или нескольких). В разработанных системах они частично формализованы и реализованы. Правила, основанные на интуиции, достаточно сложно формализовать, а соответственно и автоматизировать. Несмотря на то что правила обучения описаны достаточно основательно, в процессе решения задачи медицинской диагностики возникает ситуация неопределенности, связанная со спецификой заболевания, недостаточностью входной и текущей информации, неоднозначностью правил установления диагноза.

Соответственно, если для решения задач семантики правила обработки и оценки информации разработаны достаточно основательно, то благодаря им можно получить результат, который совпадает с целью исследования, а неопределенность в этом случае сводится к минимуму. В противном случае полученный результат может быть далек от оптимального, что создает ситуацию неопределенности.

Заключение

Итак, в задачах семантики неопределенность связана с неоднозначностью результата, полученного по смоделированной целевой функции; с выбором способа оценки точности работы определенного алгоритма; особой структурой множества комбинаторных конфигураций, которые являются аргументом целевой функции; неполной входной и текущей информацией; с многокритериальной оптимизацией; а также с нечетко разработанными правилами обработки и оценки информации. Рассмотренные ситуации неопределенности не являются полными.

Для выхода из ситуации неопределенности, которая возникает вследствие особой структуры множества комбинаторных конфигураций, оценка результата проводится как по одной, так и по нескольким целевым функциям, а также решается введением в процессе решения задачи переменных критериев. Благодаря им генерируется дополнительная текущая информация, которая влияет на прогнозирование будущих результатов. Это дает возможность получить оптимальное решение, совпадающее с целью исследования.

Для разрешения ситуации неопределенности используются самонастраивающиеся алгоритмы. Благодаря этим алгоритмам при неполной входной и текущей информации, а также в ситуации, связанной с неоднозначностью при выборе оптимального решения по нескольким критериям, находится оптимальное решение, которое удовлетворяет цели исследования. Для эффективного использования линейной свертки в многокритериальной оптимизации находится частичное решение, для которого вычисляется частичная целевая функция. Если с применением очередного критерия возникает ситуация неопределенности, вводятся дополнительные переменные критерии, которые используются как один, так и много раз в итерационном режиме.

Н.К. Тимофієва

СИТУАЦІЯ НЕВИЗНАЧЕНОСТІ, ЯКА ВИНИКАЄ В ЗАДАЧАХ СЕМАНТИКИ, ТА СПОСОБИ ЇЇ ВИРІШЕННЯ

Розглядаються різні види невизначеності, які з'являються при розв'язанні задач семантики. В теорії прийняття рішень досліджують ситуацію невизначеності, пов'язану з неповною вхідною, поточною та нечіткою інформацією. Але невизначеність в задачах семантики має інші прояви. Вирішення її проводиться різними способами в залежності від виду невизначеності. Задачі цього класу відносяться до розпізнавання, і при встановленні суті певних об'єктів вводяться міри подібності, які є суб'єктивною оцінкою. Для різних мір значення цільових функцій може відрізнятися внаслідок неоднозначності результату, одержаного за цими функціями або вибраною мірою подібності, та не задовольняти меті дослідження. При виборі результату виникає ситуація невизначеності. Але за деякими мірами подібності можна знайти і глобальний розв'язок. Такі задачі виділяються в підкласи

 ISSN 1028-0979

розв’язних задач. Оскільки задачі семантики зводяться до задач комбінаторної оптимізації, аргументом цільової функції в яких є комбінаторні конфігурації, то ситуація невизначеності може бути пов’язана з особливою структурою множини комбінаторних конфігурацій. Для її вирішення необхідно вводити кілька цільових функцій або проводити оптимізацію за кількома критеріями, які зводяться до зваженого критерію (лінійної згортки). Знаходження оптимального розв’язку проводиться самоналагоджувальними алгоритмами з урахуванням постійних та змінних критеріїв, які вводяться в процесі розв’язання задачі. Тобто в процесі роботи алгоритму генерується додаткова поточна інформація (критерії якості), яка впливає на прогнозування майбутніх результатів. Ситуація невизначеності проявляється і внаслідок нечітко розроблених правил обробки та оцінки інформації та при виборі оптимального розв’язку за кількома критеріями в багатокритеріальній оптимізації. Для виходу з цієї ситуації розробляють самоналагоджувальні алгоритми, використовують введення в процесі розв’язання задачі формальних параметрів, за допомогою яких генерується допоміжна поточна інформація, яку неможливо задати у вхідних даних. Також для вирішення ситуації невизначеності використовуються підкласи розв’язних задач, проводиться структуризація еталонної бібліотеки для зведення нерозв’язних задач до розв’язних.

Ключові слова: задачі семантики, комбінаторна оптимізація, кластеризація, самоналагоджувальні алгоритми, лінійна згортка, міри подібності, підкласи розв’язних задач.

N.K. Tymofijeva

THE SITUATION OF UNCERTAINTY THAT ARISES IN THE PROBLEMS OF SEMANTICS AND WAYS TO SOLVE IT

Various types of uncertainties that arise when solving semantics problems are considered. Decision theory investigates this situation involving incomplete input, current, and fuzzy information. But uncertainty in the problems of semantics has other manifestations. Its solution is carried out in different ways depending on its types. The problems of this class are related to recognition and when establishing the essence of certain objects, measures of similarity are introduced, which are a subjective assessment. For different measures, the values of the objective functions may differ due to the ambiguity of the result obtained for these functions or the chosen degree of similarity measures, and may not satisfy the purpose of the study. When choosing the result there is a situation of uncertainty. But with some measures of similarity, you can find a global solution. Such problems are divided into subclasses of solvable problems. Since the problems of semantics are reduced to combinatorial optimization problems, in which the argument of the objective function is combinatorial configurations, the situation of uncertainty may be related to the special structure of the set of combinatorial configurations. To solve it, it is necessary to enter several objective functions or to conduct optimization according to several criteria, which are reduced to a weighted criterion (linear convolution). Finding the optimal solution is carried out by self-tuning algorithms taking into account the constant and variable criteria, which are introduced in the process of solving the problem. That is, in the process of the algorithm generates additional current information (quality criteria), which affects the prediction of future results. The situation of uncertainty is manifested both due to developed fuzzy rules of information processing and evaluation and ambiguity in the choice of the optimal solution for several criteria in multicriteria optimization. To get out of this situation, self-tuning algorithms are developed, using the introduction of formal parameters in the process of solving the problem, which generates auxiliary current information that can not be specified in the input data. Also, subclasses of solvable problems are used to solve the situation of uncertainty, the reference library is structured to reduce unsolvable problems to solvable ones.

Keywords: semantics problems, combinatorial optimization, clustering, self-tuning algorithms, linear convolution, measures of similarity, subclasses of solvable problems.

1. Лабетова В.М. Категорія визначеності/невизначеності у світлі функціонального підходу до вивчення мови. Лінгвістичні студії. Зб. наук. праць. Вінниця : ДонНУ імені Василя Стуса, 2016. Вип. 32. С. 13–16.
2. Иваненко В.И., Лабковский В.А. Проблема неопределенности в задачах принятия решений. К. : Наук. думка, 1990. 136 с.
3. Орловский С.А. Проблемы принятия решений при нечеткой исходной информации. М. : Наука, 1981. 208 с.
4. Зайченко Ю.П. Оптимизация инвестиционного портфеля в условиях неопределенности. Состояние, проблемы, перспективы. Обчислювальний інтелект (результати, проблеми, перспективи). *Матеріали Першої Міжнародної науково-технічної конференції «Обчислювальний інтелект (ОІ-2011)»*. Україна : Черкаси, 10–13 травня 2011. С. 33–34.
5. Губарев В.Ф. Особенности и взаимосвязь задач идентификации и управления в условиях неопределенности. *Международный научно-технический журнал «Проблемы управления и информатики»*. 2010. № 1. С. 50–62.
6. Тимофієва Н.К. Про розв'язання задач комбінаторної оптимізації в умовах невизначеності. *Вісник Вінницького політехнічного інституту.* 2012. № 6. С. 157–162.
7. Тимофієва Н.К. Про деякі підходи до оцінки оптимального розв'язку задач комбінаторної оптимізації. *Системи керування та комп'ютери* (*УСіМ, Control systems & computers*). 2019. № 3 (281). С. 3–13.
8. Тимофієва Н.К. Теоретико-числові методи розв'язання задач комбінаторної оптимізації. Дисертація на здобуття наукового ступеня доктора технічних наук за спеціальністю 01.05.02 — математичне моделювання та обчислювальні методи. Рукопис. Київ : ІК ім. В.М. Глушкова НАН України, 2007. 374 с.
9. Тимофеева Н.К. О природе неопределенности и переменных критериях в задачах разбиения. *Международный научно-технический журнал «Проблемы управления и информатики».* 2009. № 5. С. 88–99.
10. Тимофієва Н.К. Самоналагоджувальні алгоритми знаходження невизначених параметрів у задачах комбінаторної оптимізації. *УСиМ.* 2009. № 4. С. 43–50.
11. Юревич Е.И. Теория автоматического управления. СПб. : БХВ Петербург, 2007. 560 с.
12. Козлов Ю.М., Юсупов Р.М. Беспоисковые самонастраивающиеся системы. М. : Наука, 1969. 455 с.

Получено 01.09.2021