

БАЗА ЗНАНИЙ ОБУЧАЮЩЕЙ СИСТЕМЫ НА ОСНОВЕ СЕМАНТИЧЕСКОЙ СЕТИ

БОСКЕБЕЕВ К.ДЖ.
izvestiya@ktu.aknet.kg

Предложенная модель позволяет формализовать базу знаний обучающей системы.

Цель исследования. Теоретическое обоснование модели представляется в виде ориентированного графа, вершины которого соответствуют понятиям предметной области изучаемой дисциплины, а дуги задают отношение «определяемое понятие – определяющее понятие» между ними. В статье рассматривается одна из актуальных граней этой проблемы – контроль усвоения субъектом обучения понятий предметной области изучаемой дисциплины, а также взаимосвязей этих понятий.

Модель исследования. В статье полагается, что база знаний обучающей системы построена на основе семантической сети, содержащей понятия предметной области изучаемой дисциплины и отношения между этими понятиями.

Стратегия диагностики, используемая в обучающей системе, зависит прежде всего от целей обучения, которые реализует эта система. Среди традиционных целей обучения выделяются следующие цели: приобретение знаний; приобретение навыков и приобретение умений [2]. Данная статья ориентирована на достижение первой из этих целей и связана с организацией контроля знаний в обучающей системе.

Будем обозначать модули рассматриваемой библиотеки знаний $m_i, i = 1, 2, \dots$. Назовем входным понятием (input concept) модуля m_i понятие $\overline{c}_{i,j}$, определение которого дано в некотором другом модуле библиотеки знаний L или иной библиотеке знаний. Набор входных понятий модуля m_i обозначим $\overline{C}_i = \{\overline{c}_{i,j}, j \in [1, \overline{n}_i]\}$, где $\overline{n}_i \geq 0$ – общее количество входных понятий. Из набора \overline{C}_i выделим входные понятия $\overline{C}I_i$, определенные в данной библиотеке знаний L , и входные понятия $\overline{C}E_i$, определение которых содержится в других библиотеках. Таким образом $\overline{C}_i = \overline{C}I_i \cup \overline{C}E_i$. Отметим, что одно или оба из множеств $\overline{C}I_i, \overline{C}E_i$ могут быть пустыми.

Аналогично назовем выходным понятием (output concept) модуля m_i понятие $c_{i,j}$, определение которого дано в данном модуле m_i . Набор выходных понятий модуля m_i обозначим $C_i = \{c_{i,j}, j \in [1, n_i]\}$, где $n_i \geq 0$ – общее количество выходных понятий.

Каждое из понятий $c_{i,j}, j \in [1, n_i]$ определяется через одно или несколько понятий из наборов \overline{C}_i, C_i . Обозначим $\overline{C}_{ij} = \{\overline{c}_{i,j,k}, k \in [1: \overline{n}_{i,j}]\}$ совокупность понятий набора, которые используются при определении понятия $c_{i,j}$. Аналогично обозначим $C_{ij} = \{c_{i,j,k}, k \in [1: n_{i,j}], k \neq j\}$ понятия набора C_i , которые используются при определении понятия $c_{i,j}$. Здесь $\overline{n}_{i,j} \geq 0, n_{i,j} \geq 0$, – количество таких понятий. Таким образом, будем полагать, что понятие $c_{i,j}$ определяется с помощью множества понятий $\overline{C}_{ij} \cup C_{ij}$. Одно или оба из множеств $\overline{C}_{ij}, C_{ij}$ могут быть пустыми. Отметим, что ситуация $\overline{C}_{ij} = 0, C_{ij} = 0$ означает, что понятие $c_{i,j}$ определяется без привлечения других понятий.

Понятия из наборов $\overline{C}_{ij}, C_{ij}$ будем называть информационно связанными (в узком смысле) с понятием $c_{i,j}$. Если понятие $c_{i,j}$ информационно связано с понятием $c_{i,k}$, и это понятие информационно связано с понятием $c_{i,l}$, то будем говорить, что понятия $c_{i,j}, c_{i,l}$ информационно связаны в широком смысле. Если понятие $c_{i,k}$, определенное в модуле m_i , информационно

связано с понятием $\bar{c}_{l,j}$, которое является входным понятием модуля m_1 и одновременно выходным понятием модуля m_2 , т.е. $\bar{c}_{l,j} = c_{2,l}$, то также будем говорить, что понятия информационно связаны в широком смысле.

Семантическую сеть $S(m_i)$ модуля m_i будем представлять в виде ориентированного графа без контуров $G(m_i)$, вершины которого соответствуют понятиям наборов \bar{C}_i, C_i а дуги отношениям «определяемое понятие – определяющее понятие» между ними. Другими словами, дуги в графе $G(m_i)$ соответствуют информационным связям понятий из наборов \bar{C}_i, C_i между собой.

Модуль m_1 использует три входных понятия $\bar{c}_{1,1}, \bar{c}_{1,2}, \bar{c}_{1,3}$ ($\bar{n}_1 = 3$), и в модуле определены четыре выходных понятия $c_{1,1}, c_{1,2}, c_{1,3}, c_{1,4}$ ($n_1 = 4$) Понятие $c_{1,4}$, к примеру, определяется с помощью двух входных понятий модуля и двух его выходных понятий: $\bar{C}_{1,4} = \{\bar{c}_{1,2}, \bar{c}_{1,3}\}$.

Аналогично информационным связям понятий определены информационные связи моделей. Модули m_i и m_j будем называть информационно связанными модулями, если хотя бы одно выходное понятие модуля m_i является входным понятием для модуля m_j или если хотя бы одно выходное понятие модуля является входным понятием m_j для модуля m_i .

Отметим следующее важное обстоятельство. В соответствии с концепцией технологии разделяемых единиц контента одно и то же понятие может быть определено в разных модулях библиотеки знаний L (в то же время ни одно из понятий не может быть определено в разных модулях учебного курса T). Назовем такие понятия кратными понятиями. Кратность понятия $c_{i,j}$ обозначим $k_{i,j} \geq 1$. Если не оговорено противное, будем далее полагать кратные понятия различными понятиями.

Современные интеллектуальные обучающие системы обеспечивают выполнение следующих правил расположения описаний понятий в модуле:

- ни одно из понятий k -го уровня в ярусно-параллельную форму (ЯПФ) модуля не может быть введено до тех пор, пока не определены все понятия всех расположенных ниже уровней ЯПФ;
- при выполнении первого правила описания понятий k -го уровня ЯПФ могут быть введены в модуле в произвольном порядке.

Наряду с этим современные обучающие системы разрешают использование в модуле понятий, которые еще не определены в данном модуле, а будут определены в нем позже. Такие понятия называются внутренними ссылочными понятиями. В терминах ЯПФ ссылка на понятие означает, что в тексте модуля при определении понятий k -го уровня используется понятие одного из расположенных выше уровней. Количество внутренних ссылочных понятий, используемых в модуле m_i , обозначается $l_i \geq 0$.

Ссылочное понятие может быть также внешним ссылочным понятием. Если в тексте некоторого модуля m_{i_j} учебного курса T используется понятие, которое определено в модуле m_{i_k} , $k \succ j$, то для модуля m_{i_j} это понятие является внешним ссылочным. Здесь принято, что если $k \succ j$, то модуль m_{i_k} в курсе T текстуально расположен позже модуля m_{i_j} .

Количество внешних ссылочных понятий модуля m_i обозначается $\bar{l}_i \geq 0$.

Пусть библиотека модулей L рассматриваемой предметной области состоит из M модулей m_i , т.е. $L = \bigcup_{i=1}^M m_i$.

Семантическую сеть $S(L)$ библиотеки L будем представлять в виде ориентированных графов $G(L)$, $G(L)$, первый из которых называется понятийным графом библиотеки L , а второй – графом информационных связей модулей этой библиотеки или ее

информационно-логическим графом. Графы $G(L)$ могут иметь контуры, количество которых обозначается $e(G(L)) = e(L)$, $e(G(L)) = e(L)$ соответственно.

Граф $G(L)$ представляет собой объединение графов семантических сетей всех модулей библиотеки L , т.е. $G(L) = \bigcup_{i=1}^M G(m_i)$.

Вершины взвешенного мультиграфа $G(L)$ соответствуют модулям библиотеки L , а дуги – информационным связям модулям между собой. В отличие от графа $G(L)$, на рисунках дуги графа изображаются жирным; рядом с дугами в качестве их веса указываются количества информационных связей между соответствующими модулями. Другими словами, если $V_{i,j}$ выходных понятий модуля m_i используются в качестве входных понятий модуля m_j , то рядом с соответствующей дугой в качестве ее веса указывается величина $V_{i,j}$.

Учебный курс, подготовленный из всех или некоторой совокупности модулей библиотеки L , обозначается T ; $T \subseteq L$. В набор входят модули $m_{i_1}, m_{i_2}, \dots, m_{i_N}$ библиотеки L , где $N \leq M$ – количество модулей в курсе T . Текстуально модули в учебном курсе T расположены именно в порядке $m_{i_1}, m_{i_2}, \dots, m_{i_N}$, т.е. первым расположен модуль m_{i_1} , вторым – модуль m_{i_2} и т.д.

Аналогично библиотеке L семантическую сеть $S(T)$ курса T будем представлять в виде ориентированных графов $G(T)$, $G(T)$, где граф $G(T)$ называется понятийным графом курса T , а граф $G(T)$ – графом информационных связей модулей этого курса, или, другими словами, его информационно-логическим графом. Поскольку допускаются контуры в графах $G(L)$, $G(L)$, графы $G(T)$, $G(T)$ также могут иметь контуры, количество которых обозначается $e(G(T)) = e(T)$, $e(G(T)) = e(T)$ соответственно.

Граф $G(T)$ представляет собой объединение графов семантических сетей всех модулей библиотеки L , входящих в учебный курс T .

Во взвешенном мультиграфе вершины соответствуют модулям $m_{i_1}, m_{i_2}, \dots, m_{i_N}$, а дуги – информационным связям модулям между собой. Аналогично графу $G(L)$ дуги графа $G(T)$ изображаются на рисунках жирным, а рядом с дугами в качестве их веса указывается кратность дуг.

Контроль знаний в обучающей системе является многоплановой проблемой.

Для решения задачи контроля понятийных знаний можно использовать тестовую подсистему обучающей системы или в рамках этой системы разрабатывать подсистему автоматического контроля понятийных знаний.

Во всех случаях для оценки уровня усвоения понятийных знаний требуется оценка сложности понятий, модулей, репозитариев (библиотек) модулей и обучающих курсов. Эти же оценки необходимы для формирования индивидуальной образовательной траектории в интеллектуальной обучающей системе, менеджмента качества учебного процесса, проектирования учебных планов образовательных программ.

Организация в обучающей системе контроля понятийных знаний субъекта обучения может требовать использования расширенной семантической сети [1], которую следует отнести к неоднородным и, возможно, N -арным семантическим сетям. К необходимости построения такой сети приводит также задача планирования в обучающей системе индивидуальной траектории обучения. В качестве меры сложности такой сети и ее фрагментов могут использоваться некоторые из мер, рассмотренных в данной работе. Систематическое рассмотрение мер сложности расширенной семантической сети, полагается, составит предмет самостоятельной публикации.

Выводы:

- Задачу оценки сложности учебного материала следует рассматривать в контексте более широкой задачи оценки качества этого материала. Для вычисления значений этой меры

предлагается классический подход Кода Шеннона: в качестве меры разнообразия принимается логарифм числа элементов, составляющих учебный материал.

- В разработке обучающей экспертной системы с узлами семантической сети (понятиями) должно связывается картинка, видеоизображение, аудиосегмент или текст и рассматриваются некоторые меры этих сущностей. В качестве одной из таких мер предлагается использовать меру количества информации по Колмогорову А.Н., заключенной в этой сущности. Значение указанной меры можно оценить размером соответствующего файла, сжатого с помощью современных алгоритмов сжатия.
- Предлагается оценивать качество семантической сети с помощью следующих критериев: достоверность (степень безошибочности данных); кумулятивность (свойство данных небольшого объема достаточно полно отражать соответствующую предметную область); непротиворечивость (отсутствие взаимоисключающих понятий).
- Предлагается в качестве меры качества учебного материала использовать его объем, а также нечеткие экспертные оценки сложности этого материала.

В данной статье меры сложности строятся на основе таких параметров семантической сети, как количество входных и выходных понятий, реберная плотность и диаметр графа, соответствующего этой сети. Предложенная модель более строго, такая сеть классифицируется как однородная (с единственным типом отношений между понятиями) и бинарная (отношения связывают только по два понятия).

Литература

1. *Федотов И.Е.* Некоторые приемы параллельного программирования: Учебное пособие. – М.: МГИРЭА(ТУ), 2008. – 188 с.
2. *Креативная педагогика: методология, теория, практика/Под ред. Ю.Г.Круглова.* –М.:МГОПУ им. М.А.Шолохова, «Альфа», 2002. –240 с.