УДК.:004.413.5:004.822

МОДЕЛЬ ОЦЕНКИ СЛОЖНОСТИ СЕМАНТИЧЕСКОЙ СЕТИ В ИНТЕЛЛЕКТУАЛЬНЫХ ОБУЧАЮЩИХ СИСТЕМАХ

БОСКЕБЕЕВ К.ДЖ., МАМБЕТОВ Н.Ж.
izvestiya@ktu.aknet.kg

В статье предложена формальная мера оценки сложности семантической сети в обучающей системе

В данной работе меры сложности строятся на основе таких параметров семантической сети, как количество входных и выходных понятий, реберная плотность и диаметр графа, соответствующего этой сети.

Организация в обучающей системе контроля понятийных знаний субъекта обучения может требовать использования расширенной семантической сети, 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. Формализация сложности семантической сети приведена ниже.

Будем обозначать модули рассматриваемой библиотеки знаний $m_i, i = 1,2,...$, 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 $\overline{CI_i}$, определенные в данной библиотеке знаний L, и входные понятия $\overline{CE}_i$, определение которых содержится в других библиотеках. Таким образом $\overline{C}_i = \overline{CI}_i \cup \overline{CE}_i$. Отметим, что одно или оба из множеств $\overline{CI}_i$, $\overline{CE}_i$ 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 . 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.

ABBYY PDF Transformer 2.0 Click here to buy www.ABBYY.com

Понятия из наборов $\overline{C}_{ij}$, C_{ij} будем называть информационно связанными (в узком смысле) с понятием $c_{i,j}$. Если понятие информационно связано с понятием $c_{i,l}$, то будем говорить, что понятия $c_{i,j}$, $c_{i,l}$ информационно связаны в широком смысле [2]. Если понятие $C_{1,k}$, определенное в модуле m_1, информационно связано с понятием $\overline{c}_{1,j}$, которое является входным понятием модуля m_1 и, одновременно, выходным понятием модуля m_2, т.е. $c_{1,j} = c_{2,1}$, то также будем говорить, что понятия $c_{1,k}, c_{2,1}$ 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 $\overline{C_i}$, C_i, а дуги - отношениям «определяемое понятие – определяющее понятие» между ними. Другими словами, дуги в графе соответствуют информационным связям понятий из наборов между собой.

Введенные обозначения иллюстрирует рис. 1. Модуль m_1 на этом рисунке использует три входных понятия $\overline{c}_{1.1}, \overline{c}_{1.2}$, $\overline{c}_{1.3}$, ($\overline{n}_1 = 3$), и в модуле определены четыре выходных понятия $\overline{c}_{1.1}, \overline{c}_{1.2}$, $\overline{c}_{1.3}$, ($n_1 = 4$). Понятие , к примеру, определяется с помощью двух входных понятий модуля и двух его выходных понятий: $\overline{C}_{1,4} = \{\overline{c}_{1,2}, \overline{c}_{1,3}\}$; $C_{1,4} = \{c_{1,2}, c_{1,3}\}$.

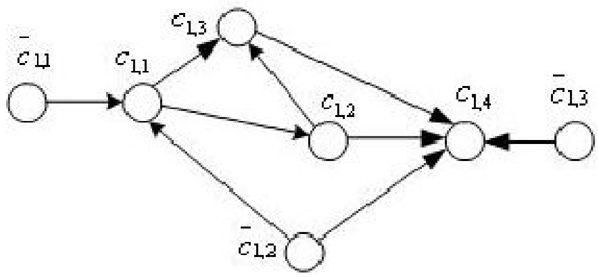

Рис. 1. Пример графа семантической сети модуля m_1.

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 m_j является входным понятием 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). 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 ярусно-параллельной формы (ЯПФ) модуля не может быть введено до тех пор, пока не определены все понятия всех расположенных ниже уровней ЯПФ [1];
- 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 будет 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.

Ссылочное понятие может быть также внешним ссылочным понятием. Если в тексте некоторого модуля учебного курса используется понятие, которое определено в модуле $m_{i_k}, k > j$, то для модуля это понятие является внешним ссылочным. Здесь принято, что если $k > j$, то модуль m_{i_k} в курсе текстуально расположен позже модуля. 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. $L = \bigcup_{i-1}^{M} m_i$ [4].

Семантическую сеть [2] $S(L)$ 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 $G(L)$ 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 : $T \subseteq L$. В набор входят модули m_i, ,…, библиотеки L, где $N \leq M$ - количество модулей в курсе . Текстуально модули в учебном курсе расположены именно в порядке m_i, ,…,, т.е. первым расположен модуль m_i, вторым – модуль m_{i+1} и т.д.

Аналогично библиотеке L, семантическую сеть S(T) курса будем представлять в виде ориентированных графов, $G(T)$, где граф называется понятийным графом курса, 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 также могут иметь контуры, количество которых обозначается $e(G(T)) = e(T), e(G(T)) = e(T)$, соответственно.

Граф представляет собой объединение графов семантических сетей всех модулей библиотеки L, входящих в учебный курс T, т.е. $G(T) = \bigcup_{j \in [tN]} G(m_i)$

Во взвешенном мультиграфе $G(T)$ вершины соответствуют модулям $m_{i_k}, m_{i_k}, \ldots, m_{i_k}$, 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.

Высота графа. Рассмотрим ярусно-параллельную форму (ЯПФ) ориентированного графа без контуров G [2]. ЯПФ орграфа G строится по следующему алгоритму:

- на первый ярус ЯПФ помещаются все те вершины графа G, в которые не входят дуги от других вершин графа;
- на k-й уровень помещаются те вершины графа G, в которые входят дуги только от его вершин, расположенных на предыдущих (k-1) ярусах; $k \geq 1$.

Проиллюстрируем этот алгоритм на примере графа $G(m_i)$. Если множество $\overline{C}_i$ не пусто, то на первом ярусе ЯПФ этого графа размещаются входные понятия $\overline{C}_i$ модуля ; m_i на втором ярусе – понятия, определяемые только с помощью понятий $\overline{C}_i$; на третьем ярусе – понятия, определяемые только с помощью понятий $\overline{C}_i$ и понятий второго яруса и т.д. (см. рис. 2). Если модуль m_i не использует входных понятий, т.е. если множество $\overline{C}_i$ пусто, то на первом ярусе соответствующей ЯПФ располагаются выходные понятия этого модуля, которые в своих определениях не содержат других понятий.

Номер яруса ЯПФ графа G, на котором находится вершина этого графа, называется высотой вершины, количество ярусов в ЯПФ графа G называется высотой ЯПФ этого графа (см. рис. 2).

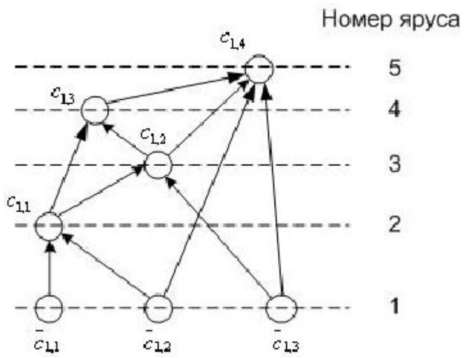

Рис. 2. Пример. Граф в ярусно-параллельной форме

Высоту вершины $c_{i,j}$ в графе $G(m_i)$ будем называть высотой понятия $c_{i,j}$ и обозначать $h(c_{i,j})$. Высота понятия зависит от контекста, в котором рассматривается данное понятие. Поясним данное утверждение примером. Пусть учебный курс T состоит из модулей m_1, m_2, и граф $G(m_1)$ первого из них в ярусно-параллельной форме представлен на Рис. 3, а граф $G(m_2)$ второго, также в ярусно-параллельной форме, – на рис. 4. Положим, что в качестве входных понятий модуля m_1 используются следующие понятия модуля m_2: $\bar{c}_{1,1} - \bar{c}_{2,1}; \bar{c}_{1,2} - \bar{c}_{1,3} - c_{2,2}$ Тогда в ярусно-параллельной форме понятийный граф $G(T)$ учебного курса T будет иметь вид, представленный на рис. 4 (а). Из этого рисунка следует, что высота понятия $c_{1,3}$ в нем равна 6.

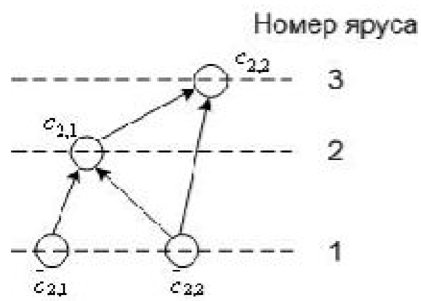

Рис. 3. Пример. Граф $G(m_2)$ модуля m_2 в ярусно-параллельной форме.

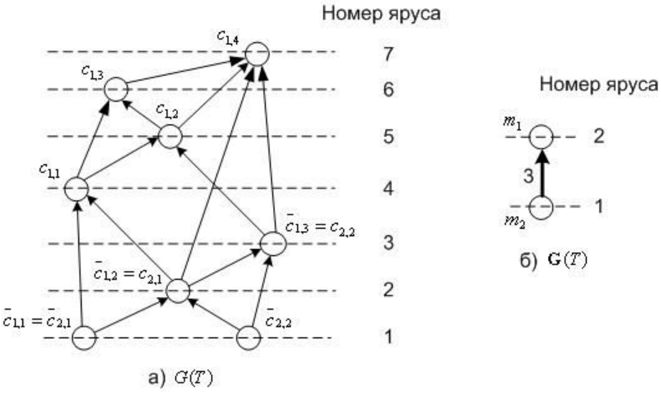

Рис. 4. Графы $G(T)$ учебного курса $T = m_1 \bigcup m_2$

ABBYY PDF Transformer 2.0 Click here to buy www.ABBYY.com

Контекст, в котором рассматривается высота понятия, будем обозначать надиндексом, так что $h(c_{i,j}), h^L(c_{i,j}), h^T(c_{i,j})$, - высоты понятия $c_{i,j}$, определенного в модуле m_i, в библиотеке L и учебном курсе T, соответственно. В последнем случае полагается, что модуль m_i входит в число модулей курса T. Высоту ЯПФ графа $G(m_i)$ будем называть высотой модуля m_i и обозначать $h(G(m_i)) = h(m_i)$. По аналогии определяются высоты библиотеки $h(G(L)) = h(L)$, $h(G(L)) = h(L)$ и высоты учебного курса $h(G(T)) = h(T)$, $h(G(T)) = h(T)$. Таким образом, $h(L)$, $h(T)$ - высоты библиотеки L и курса T для соответствующих понятийных графов, а $h(L)$, $h(T)$ - те же высоты, но для графов информационных связей модулей.

В случае, если граф G имеет контуры, применение рассмотренного алгоритма построения его ЯПФ приводит к тому, что, начиная с некоторого уровня, оказывается невозможным ни одну из оставшихся вершин отнести к следующему уровню. При обнаружении этого факта необходимо разорвать одну или несколько из дуг, соединяющих оставшиеся вершины.

ВЫВОДЫ

1. Предложенные в работе некоторые из мер сложности могут быть использованы в качестве мер усвояемости учебного материала и мер его адаптивности. Это обстоятельство позволяет формально ставить задачу оценки качества учебных материалов, как трехкритериальную задачу (с критериями: сложность; усвояемость; адаптивность). Результаты работы позволяют значения всех трех критериев качества вычислить автоматически только на основе анализа семантической сети соответствующего учебного материала.
2. Предлагаются и другие меры, например: меры, использующие объемы модулей, библиотек и учебных курсов; меры, использующие метаданные тех же единиц контента и др. С нашей точки зрения, меры такого сорта менее содержательны и объективны, чем меры на основе метрик семантической сети.

Литература

1. Федотов И.Э. Некоторые приемы параллельного программирования: Учебное пособие. - М.: Изд-во.МГИРЭА(ТУ), 2008. – 188 с.

2. Соломатин Н.М. Перспективы развития вычислительной техники: В 11 кн. Справ. Пособие / Под ред. Ю.М. Смирнова. Кн. 1: Информационные семантические системы. - М.: Высш. Шк., 1989. – 127 с