$$\Delta \geqslant f(\tau) - f(\tau^2) \geqslant 4 \sum_{k=1}^{n/2} \sqrt{m^k} \sum_{i=1}^{[(n-2)/4]} \sum_{j=i+1}^{n/2-i} \sqrt{m^j} \cos \frac{2\pi i}{n} +$$

$$+ 4 \Big(\sum_{i=1}^{[(n-2)/4]} \sum_{j=i+1}^{n/2-i} \sqrt{m^j} \cos \frac{2\pi i}{n} \Big)^2.$$

Сравнивая оценку, полученную в теореме, с последней, приходим к следующим выводам:

1) если $\sqrt{m^1} + \sqrt{m^{n/2}} \leqslant \sum_{i=2}^{[(n-2)/4]} \sum_{j=i+1}^{n/2-i} \sqrt{m^j} \cos \frac{2\pi i}{n}$, то алгоритм 2 строит решение, которое асимптотически в 9 раз лучше по значению целевой функции, чем самое «плохое решение».

2) если $m^i = m^j$ для всех i и j, $1 \leqslant i, j \leqslant n/2$, то алгоритм 2 строит перестановку, на которой значение $f(\tau^2)$ асимптотически в $0(n^3)$ раз лучше, чем на самой «плохой перестановке».

3) если $m^i = 0$ для всех i, $1 \leqslant i \leqslant n/2$, то алгоритм 2 строит оптимальное решение.

Поступила в редакцию 16.10.87.

УДК 002.513.5

С. Ф. ЛИПНИЦКИЙ

МАТЕМАТИЧЕСКАЯ МОДЕЛЬ СИНТАКСИЧЕСКОГО АНАЛИЗА ВХОДНЫХ СООБЩЕНИЙ В ИНТЕЛЛЕКТУАЛЬНОЙ ИНФОРМАЦИОННОЙ СИСТЕМЕ

Характерной чертой развития теории и практики современных автоматизированных информационных систем (АИС) является повышение их интеллектуальных возможностей в направлении совершенствования механизмов общения пользователей с АИС. Один из путей интеллектуализации АИС состоит в использовании в них входных языков, приближающихся по своей структуре и семантической силе к естественному языку [1, 2]. Обработка входных текстов в этих случаях осуществляется в несколько этапов, одним из которых является синтаксический анализ. Конечная цель синтаксического анализа — построение синтаксического графа каждого предложения текста.

В данной статье предлагаются математическая модель и алгоритм синтаксического анализа. В отличие от существующих предлагаемый алгоритм построен на основе изучения в рамках модели основных синтагматических отношений на цепочках входного языка, что позволило существенно упростить и ускорить процедуру анализа.

Свойства синтагматических отношений. Пусть U — словарь входного языка (входной словарь), элементы которого будем называть словами, U^* — множество всех цепочек в словаре U, а U^+ — множество всех непустых цепочек в U. По аналогии с [3] определим на множестве U^+ отношения парадигматического и синтагматического подчинения и синтагматической эквивалентности.

Определение 1. Рефлексивное и антисимметричное отношение Δ на множестве U^+ назовем отношением парадигматического подчинения, если для любых цепочек $a, b, c \in U^+$ и $m, n \in U^*$:

1) $(mbn, c) \in \Delta$ и $(a, b) \in$, то $(man, c) \in \Delta$;

2) $(a, mbn) \in \Delta$ и $(b, c) \in \Delta$, то $(a, mcn) \in \Delta$;

3) $(a, b) \in \Delta$ и $(c, b) \in \Delta$, то существует цепочка $d \in U^*$ такая, что $(d, a) \in \Delta$ и $(d, c) \in \Delta$.

Определение 2. Отношение δ на множестве U^+ назовем отношением синтагматического подчинения, если $(a, b) \in \delta$ тогда и только тогда, когда:

1) существуют цепочки m, $n \in U^+$ такие, что $(m, a) \in \Delta$ и $(n, b) \in \Delta$;

2) для m и n, удовлетворяющих соотношениям п. 1, справедливо $(m, mn) \in \Delta$.

Цепочки вида ab, где $(a, b) \in \delta$ или $(b, a) \in \delta$, будем называть синтагматическими структурами, а при $a, b \in U$ — синтагмами. Если $(a, b) \in \delta$, то будем говорить, что a — определяемый, а b — определяющий члены структур ab и ba.

Определение 3. Отношение ω на множестве U^+ назовем отношением синтагматической эквивалентности, если $(a, b) \in \omega$ тогда и только тогда, когда существует цепочка $l \in U^+$ такая, что $(l, a) \in \Delta$, $(l, b) \in \Delta$, $(l, ab) \in \Delta$ и $(l, ba) \in \Delta$.

Если $(a_i, a_{i+1}) \in \omega$ $(i = \overline{1, n-1})$, то цепочку $a_1 a_2 \ldots a_n$ назовем композицией цепочек $a_1, a_2, \ldots, a_n$.

Определение 4. 1) Слово a входного словаря назовем входным сообщением, если существует цепочка $b \in U^+$ такая, что $(a, b) \in \delta$.

2) Если a и b — входные сообщения, то цепочки ab и ba такие, что $(a, b) \in \delta$ или $(a, b) \in \omega$, будем называть входными сообщениями.

Теорема 1. 1) Если цепочка a является конкатенацией цепочек c и d (в любом порядке) и $(c, d) \in \delta$, то $(a, b) \in \delta$ тогда и только тогда, когда $(c, b) \in \delta$.

2) Если цепочка b является конкатенацией цепочек c и d (в любом порядке) и $(c, d) \in \delta$, то $(a, b) \in \delta$ в том и только в том случае, когда $(a, c) \in \delta$.

Доказательство. *Необходимость.* Докажем п. 1; схема доказательства п. 2 аналогична. В силу определения 2 существуют цепочки $m_i, n_i \in U^+$ $(i = 1, 2)$ такие, что $(m_1, c) \in \Delta$, $(n_1, d) \in \Delta$, $(m_1, m_1 n_1) \in \Delta$ и $(m_2, a) \in \Delta$, $(n_2, b) \in \Delta$, $(m_2, m_2 n_2) \in \Delta$. Тогда по определению 1 $(m_1, a) \in \Delta$ и $(m_2, a) \in \Delta$, откуда следует, что существует цепочка $m \in U^+$, для которой $(m, m_1) \in \Delta$ и $(m, m_2) \in \Delta$. С учетом того, что $(m, a) \in \Delta$ и $(n_2, b) \in \Delta$, имеем $(m, mn_2) \in \Delta$, т. е. при $(m, c) \in \Delta$ получаем $(c, b) \in \Delta$.

Достаточность. Ограничимся доказательством п. 2; п. 1 доказывается аналогично. Согласно определению 2, существуют цепочки $m_i, n_i \in U^+$ $(i = 1, 2)$ такие, что $(m_1, a) \in \Delta$, $(n_1, c) \in \Delta$, $(m_2, c) \in \Delta$ и $(n_2, d) \in \Delta$, откуда в силу определения 1 существует цепочка $l \in U^+$, для которой $(l, n_1) \in \Delta$, $(l, m_2) \in \Delta$. Тогда $(l, c) \in \Delta$, $(m_1, m_1 l) \in \Delta$. С учетом того, что $(l, cd) \in \Delta$, имеем $(a, b) \in \delta$.

Аналогичные утверждения нетрудно доказать и для цепочек, состоящих в отношениях синтагматических эквивалентности и подчинения.

Теорема 2. 1) Если $a = cd$ и $(a, b) \in \delta$, а $(c, d) \in \omega$, то $(c, b) \in \delta$ и $(d, b) \in \delta$.

2) Если $a = cd$ и $(c, d) \in \omega$, $(d, b) \in \delta$, то $(a, b) \in \delta$.

3) Если $b = cd$ и $(a, b) \in \delta$, $(c, d) \in \omega$, то $(a, c) \in \delta$ и $(a, d) \in \delta$.

4) Если $b = cd$ и $(c, d) \in \omega$, $(a, c) \in \delta$, то $(a, b) \in \delta$.

Теоремы 1 и 2 позволяют свести процесс синтаксического анализа входного сообщения к анализу его синтагм и синтагм цепочек, получаемых исключением из сообщения определяющих членов синтагматических структур. Пример анализа сообщения: ЯЗЫКИ ПРЕДСТАВЛЕНИЯ ЗНАНИЙ В ИНТЕЛЛЕКТУАЛЬНЫХ ИНФОРМАЦИОННЫХ СИСТЕМАХ РАЗВИВАЮТСЯ И СОВЕРШЕНСТВУЮТСЯ БЫСТРЫМИ ТЕМПАМИ поэтапно представлен в таблице.

На первом этапе во входном сообщении выявляются все синтагмы входного сообщения.

На втором и последующих этапах отыскиваются все синтагмы в подцепочках входного сообщения, полученных путем последовательного исключения определяющих членов синтагм, выявленных на предыдущих этапах анализа.

Признаком завершения синтаксического анализа является выделение последней синтагматической структуры (этап 7).

Пример синтаксического анализа входного сообщения

Этапы синтаксического анализа				
1	2	3	. . .	7
языки представления знаний в интеллектуальных информационных системах развиваются и совершенствуются быстрыми темпами	языки представления знаний в интеллектуальных системах развиваются совершенствуются темпами	языки представления знаний в системах совершенствуются		языки совершенствуются

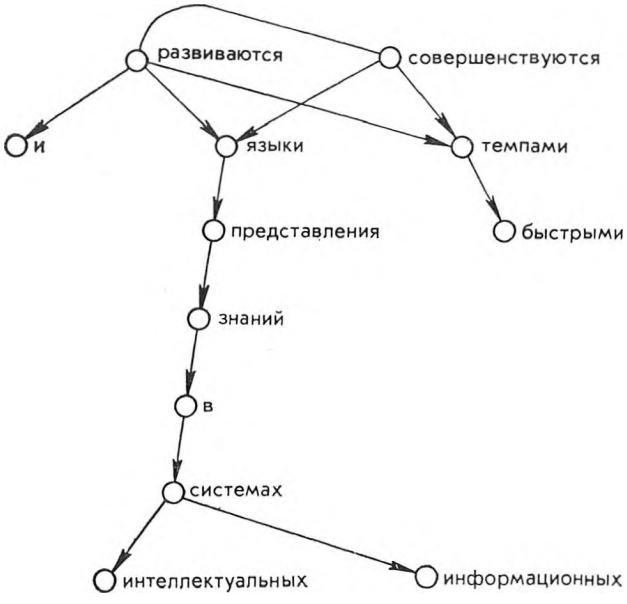
Пусть V — словарь, элементы которого будем называть словами синтаксического языка, а α: $U \to V$ — инъективное отображение. Каждое слово синтаксического языка состоит из грамматического кода, указывающего на часть речи и морфологические значения, символа синтаксического значения и основы соответствующего слова входного сообщения [3].

Определение 5. Смешанный граф $(A\alpha, R \cup D)$, где A — множество всех слов некоторого входного сообщения u, а R и D — множества ребер и дуг таких, что соответственно $(a_1\alpha, a_2\alpha) \in R$ и $(a_3\alpha, a_4\alpha) \in D$ $(a_i \in A, i = \overline{1, 4})$, если $(a_1, a_2) \in \omega$ и $(a_3, a_4) \in \delta$, будем называть синтаксическим графом сообщения u.

Пример синтаксического графа упомянутого выше входного сообщения приведен на рисунке.

Теорема 3. 1) Для каждого входного сообщения существует единственный синтаксический граф.

2) Синтаксический граф входного сообщения является ордеревом,



Пример синтаксического графа входного сообщения

если это сообщение не содержит подцепочек, являющихся композициями цепочек во входном словаре.

Теорему 3 нетрудно доказать индукцией по числу шагов процедуры построения входного сообщения (см. определение 4).

Алгоритм синтаксического анализа. На входе алгоритма — входное сообщение, на выходе — синтаксический граф в виде совокупности синтагм и композиций слов. Алгоритм включает следующие шаги.

1. Провести морфологический анализ входного сообщения [3], т. е. для каждой его словоформы выявить морфологические признаки (падеж, число, род) и принадлежность ее к некоторой лексической категории (части речи). Если сообщение состоит из одного слова, то — КОНЕЦ.

2. Найти в сообщении все композиции слов, сформировать соответствующие им части синтаксического графа и оставить в каждой композиции по одному слову, исключив остальные.

3. Найти в сообщении все синтагмы и сформировать соответствующие им части синтаксического графа. Если в сообщении образовалась единственная синтагматическая структура, то перейти к п. 14.

4. Исключить из всех синтагматических структур вида $a_1a_2 \ldots a_n$, где $n=1, 2$ $(a_{i+1}, a_i) \in \delta$ $(i=\overline{1, n-1})$, по одному слову, являющемуся определяющим членом синтагмы.

5. Найти в сообщении все появившиеся после исключения определяющих членов синтагм (см. п. 4) композиции слов, сформировать соответствующие им части синтаксического графа и оставить в каждой композиции по одному слову.

6. Найти в сообщении все синтагмы, появившиеся после исключения указанных в п. 4 слов, и сформировать соответствующие им части синтаксического графа.

7. Если все слова — определяющие члены синтагм — исключены, то перейти к п. 8, иначе — к п. 4.

8. Восстановить все синтагматические структуры, определяемые члены которых не вошли в состав других синтагматических структур.

9. Исключить из всех синтагматических структур вида $b_1b_2 \ldots b_n$, где $(b_i, b_{i+1}) \in \delta$ $(i=\overline{1, n-1})$, по одному слову, являющемуся определяющим членом синтагмы.

10. Найти в сообщении все появившиеся после исключения определяющих членов синтагм (см. п. 9) композиции слов, сформировать соответствующие им части синтаксического графа и оставить в каждой композиции по одному слову, исключив остальные.

11. Найти в сообщении все синтагмы, появившиеся после исключения слов, указанных в п. 9, и сформировать соответствующие им части синтаксического графа.

12. Если все слова — определяющие члены синтагм — исключены, то перейти к п. 13, иначе — к п. 4.

13. Если полученное сообщение не является синтагматической структурой, то — КОНЕЦ (исходная цепочка не является входным сообщением).

14. Дополнить полученный граф, если необходимо, дугами, соответствующими синтагмам, которые, согласно теореме 2, образованы словами, исключенными из композиций в пп. 5, 10. КОНЕЦ.

Очевидно, алгоритм синтаксического анализа заканчивает свою работу не более чем за $n^2/4$ проходов шагов 4—12 (n — число слов во входном сообщении).

Теорема 4. Граф, построенный в результате анализа входного сообщения в соответствии с алгоритмом, является синтаксическим графом этого сообщения.

Теорема 4 может быть доказана путем рассмотрения всех возможных случаев отклонения полученного графа от синтаксического.

Предложенная в данной статье математическая модель и алгоритм применимы при синтаксическом анализе проективных и слабо проективных предложений, характерных для научной и деловой прозы.

Список литературы

1. Гладкий А. В. Синтаксические структуры естественного языка в автоматизированных системах общения. М., 1985.
2. Попов Э. В. Общение с ЭВМ на естественном языке. М., 1982.
3. Липницкий С. Ф., Яковишин В. С. // Весці АН БССР. Сер. фіз.-тэхн. навук. 1987. № 4.

Поступила в редакцию 27.10.87.

УДК 519.8

Х. Д. ШУНГАРОВ

ОБ ЭФФЕКТИВНОСТИ ОДНОГО АЛГОРИТМА РЕШЕНИЯ ЗАДАЧИ ПОКРЫТИЯ ГРАФА ЗВЕЗДАМИ

Ранее рассматривалась трехкритериальная задача покрытия взвешенного полного графа звездами и был предложен точный алгоритм ее решения [1]. В данной работе исследуется двухкритериальная задача покрытия произвольного взвешенного графа звездами, которая состоит в следующем: задан n-вершинный неориентированный граф $G=(V, E)$, каждое ребро $e=(i, j)$ которого взвешено числом $w(e)=a_{ij}$, $a_{ij}\in \in\{1, 2, \ldots, R\}$. Напомним, что h-звездой называется полный двудольный граф $K_{1,h-1}$. Пусть $W_n=\{h : h\in N_n\}$, n кратно h, где $N_n=\{1, 2, \ldots, n\}$.

Покрытием графа $G=(V, E)$ h-звездами (h-покрытием) будем называть его остовный подграф $x=(V, E_x)$, $E_x\subseteq E$, каждая компонента связности которого представляет собой h-звезду, $h\in W_n$ [1].

На множестве $X=\{x\}$ всевозможных h-покрытий, $h\in \mathbf{W}_n$, зададим векторную целевую функцию ВЦФ $F(x)=(F_1(x), F_2(x))$, частные критерии которой $F_k(x)\to \min$, $k=1, 2$, имеют вид: $F_1(x)=\frac{1}{n-|x|}\sum\limits_{e\in E_x} w(e)$, $F_2(x)=|x|$, где $|x|$ — число звезд в покрытии x. Тем самым критерий $F_1(x)$ — удельный вес покрытия x. ВЦФ $F(x)$ определяет на X паретовское множество $\widetilde{X}\subseteq X$ или множество эффективных альтернатив [2, 3]. Задача состоит в построении алгоритма нахождения так называемого полного множества альтернатив (ПМА). ПМА определяется как такое минимальное по мощности подмножество $X^0\subseteq X$, что $F(X^0)=F(\widetilde{X})$, где $F(X^*)=\{F(x) : x\in X^*\}$, $\forall X^*\subseteq X$ [4].

Среди возможных подходов к решению этой задачи наиболее естественным представляется использование локальных алгоритмов и процедур [5].

В настоящей работе найдены ограничения на параметры задачи, при которых мощность ПМА X^0 почти всегда равна единице. При этих условиях разработан статистически эффективный алгоритм построения такого ПМА, который состоит в следующем.

Алгоритм α начинает свою работу с разбиения графа G на γ подграфов $G_k=(V_k, E_k)$, $k=\overline{1, \gamma}$, затем вычисляется число $N=\frac{n}{h}$ звезд в покрытии x, величина $\nu=\left[\frac{N}{\gamma}\right]$ и определяется мощность $|V_k|=\nu h$ для $k=\overline{1, l}$ и $|V_k|=(\nu+1)h$ для $k=\overline{l+1, \gamma}$, где $l=(\nu+1)\gamma-N$.

Алгоритм α состоит из этапов α_1 и α_2. Этап α_1 состоит из γ подэтапов α_1^k, $k=\overline{1, \gamma}$. Для $k=\overline{1, l}$ ($k=\overline{l+1, \gamma}$) подэтап α_1^k состоит из $\nu(\nu+1)$ шагов $s=\{1, \ldots, \nu\}$ ($s=\{1, \ldots, \nu+1\}$). Результатом шага s является r-звезда, покрывающая вершины множества V_k. Каждый из таких