

## ВЫВОДЫ ПО АНАЛОГИИ И МЕТОДЫ ИДЕНТИФИКАЦИИ СООБЩЕНИЙ

## ANALOGICAL INFERENCES AND METHODS OF MESSAGES' IDENTIFICATION

**Аннотация.** В статье показано, что при идентификации сообщений в интерактивных системах «человек–компьютер» особые типы аналогий могут давать заключения высокой достоверности.

**Summary.** It is shown that in the domain of messages' identification some special types of analogies can provide conclusions with high degree of credibility.

Проблема повышения достоверности идентификации и коррекции сообщений является традиционной областью приложения методов теории кодирования. Качественная коррекция сообщения достигается с помощью введения избыточности и подходящей оценки “близости” сообщений на приеме [1, 2]. В качестве таких оценок используются различные виды расстояний между кодами, причем интенсивно разрабатываются и изучаются новые меры таких расстояний [3].

Развитие средств связи и вычислительной техники вводит в круг рассмотрения задачи, в которых нарушается допущение о том, что искажения сообщения вносятся исключительно трактом передачи. Таковы задачи идентификации звуковых сигналов (в том числе распознавания речи), машинного зрения, обработки естественно-языковых запросов, компьютерного тестирования. В подобных случаях сообщение может уже на входе содержать искажения, и необходимо отнести его к некоторому заданному классу сообщений. Это достигается посредством той или иной адекватной меры расстояния (или “сходства”) между сообщениями, позволяющей с “наибольшей степенью правдоподобия” поместить сообщение в подходящий класс. Для различных типов задач разработаны меры сходства, использующие различные математические средства [4...10].

Однако расширение области исследуемых типов возможных искажений сообщений вовлекает в рассмотрение новые методы оценки их сходства. Разнообразные типы отношений сходства являются предметом изучения теории выводов по аналогии [11]. Поэтому уместно применение ряда разработанных в этой теории методов к задачам идентификации сообщений.

Целью данной работы является демонстрация эффективности некоторых типов выводов по аналогии при идентификации и коррекции сообщений в интерактивных системах.

**1. Математические и компьютерные модели выводов по аналогии.** Обычно “аналогией” называют вывод, при котором некоторая информация  $\mathfrak{Z}$  об объекте  $a$  ( $a$  называют *моделью*, источником или базой) “переносится” на другой объект  $b$  (называемый *прототипом*, целью или мишенью) на основании некоторого *сходства* или *подобия* между  $a$  и  $b$ . Виды аналогии зависят от истолкования понятий *объект*, *информация* и *сходство*.

К античности восходит понимание аналогии как вывода на основании тождества или сходства *отношений* в модели и прототипе. Впоследствии модель и прототип часто трактовались как заданные наборами *свойств*, а аналогия соответственно – как вывод на основании общности части этих свойств (этот тип вывода называют “парадегмой”).

Тезис «любая аналогия хромает» неверен. Аналогия может давать достоверные или “практически достоверные” заключения. Необходимым условием этого является алгоритмический характер *процедуры сравнения*  $S(a, b)$  модели и прототипа, приводящей к той или иной оценке их сходства. Другим условием высокой степени правдоподобия аналогии является надлежащее ограничение “фона”, или “контекста” вывода. Например, “в контексте” компьютерного тестирования знаний по математике совпадение значений функции-модели (хранимой в базе данных) и функции-прототипа (вводимой в качестве ответа на вопрос) всего для одного-двух значений их аргументов достаточно для признания ответа правильным.

Поскольку аналогия есть вывод на основании сходства, элементы аналогии присутствуют во многих методах идентификации, часто сочетаясь с индуктивными методами. К аналогии “в тесном смысле” относится разработка специальных способов формального представления модели и прототипа, ориентированных на вычисление различных мер их “частичного тождества”, “эквивалентности”, или “подобия”.

В современной логике и математике строятся различные формальные (иногда полуформальные) модели выводов по аналогии. Примеры: (I) А.Уёмов [11] строит обширную классификацию аналогий по различию структур модели  $a$ , прототипа  $b$ , переносимой информации  $\mathfrak{Z}$  и основания вывода  $S(a, b)$ ; затем формулирует для некоторых типов аналогий условия правдоподобия, а иногда и досто-

верности. (II) Часто аналогия “встраивается” в модели вероятностного логического вывода. При этом она обычно трактуется как парадейгма. Именно, вводятся такие пропозициональные вероятностные меры, которые были бы релевантны правилу: «Если объекты  $a$  и  $b$  оба имеют свойство  $\alpha$ , и, кроме того,  $a$  имеет свойство  $\beta$ , то *caeteris paribus* более вероятно, что  $b$  имеет свойство  $\beta$ , чем свойство не- $\beta$ » [12]. (III) Имеются работы, в которых на основе той или иной математической “теории объектов” вводятся особые понятия “сходства” и “аналогии”. Так, В. Булитко [13] предлагает моделировать аналогию как алгоритмическую сводимость формальной теории, описывающей прототип  $b$ , к формальной теории модели  $a$ , а типы аналогий трактуются как типы отношений сводимости. (IV) Отдельную область составляют методы поиска математических доказательств, использующие аналогии с уже имеющимися доказательствами других задач; здесь обычно используют логику второго порядка (для построения “обобщенных” доказательств) и метод резолюций [14, 15].

Когнитивная психология ставит задачу изучить выводы по аналогии в тех ее формах, которые реально используются людьми в рассуждениях, и построить близкие к этим формам компьютерные модели аналогий. Во всех таких моделях доминирующее место занимает аналогия на основе сходства отношений [16...18]. Разные модели акцентируют внимание на разных этапах вывода (таких, как *поиск* аналогов, *способы оценки* их сходства и т. п.). Другие важные различия связаны с выбором схем представления вывода: в одних моделях используются продукции, в других – сети квази-нейронов, многоагентные системы, или же рассуждения на основе прецедентов. Общей чертой является стремление максимально расширить контекст аналогии (в результате степень правдоподобия вывода в этих моделях сильно зависит от детализации описаний задач, к которым их применяют).

Модели аналогий, разрабатываемые в computer science, заимствуют многие идеи логики и когнитивной психологии. Но здесь не ставится цель дать «общую схему» аналогии. Многие исследования направлены на узкие, специализированные области применения, в результате чего становятся возможными аналогии, близкие к достоверным. Так, в роботехнике изучаются различные пространственные аналогии, способствующие ориентации робота [19, 20]. В задачах таксономии и поиска в базах знаний используются различные меры аналогичности предикатов и продукций, позволяющие существенно сокращать время поиска [21, 22]. Аналогии, основанные на алгоритмах оценок сходства последовательностей белков ДНК, с успехом применяются в молекулярной биологии [23]. Часто аналогии служат базовыми инструментами компьютерных систем обучения в той или иной “хорошо очерченной” предметной области: решение алгебраических уравнений [24]; разработка компьютерных интерфейсов [25, 26]; описание механизмов кровообращения [27]. Обзор ряда других областей приложения алгоритмов, использующих аналогии, см. в [20, 28...30].

Хотя все упомянутые алгоритмы исходят из трактовки аналогии как вывода на основании частичного сходства модели с прототипом, в любом из них конкретный *вид аналогии* определяется *алгоритмом сравнения* модели и прототипа, а *адекватность* алгоритма (для получения правдоподобных заключений) – *контекстом* вывода. Контекст связан, помимо прочего, с двумя важнейшими факторами: 1) способом представления информации о модели и прототипе; 2) переносимыми с модели на прототип признаками. В разных контекстах оправданными будут разные способы сравнения модели и прототипа, и нет оснований искать “универсально эффективную” предметно независимую схему аналогии (хотя, разумеется, разные схемы могут различаться по широте применения). Пользуясь терминологией, принятой в исследованиях по искусственному интеллекту, можно сказать, что эффективность аналогии зависит от *онтологии* [4], характеризующей предметную область вывода.

## 2. Аналогии в задачах идентификации сообщений

**2.1. «Гипотеза неустранимой неоднозначности».** Мы рассмотрим здесь один класс задач идентификации, связанный с принятием особых допущений об эквивалентности синтаксически различных сообщений, обрабатываемых (в частности) в интерактивных системах «человек–компьютер».

Обсуждаемый класс задач связан с *отказом* от допущения, что любые синтаксически различные сообщения являются семантически различными. В других терминах это отбрасываемое допущение можно сформулировать так: сообщение, поступающее на вход канала передачи, не содержит никаких ошибок, а возможные искажения на выходе возникают при передаче.

Ниже будут допускаться иные источники порождения ошибок, вносимые не в процессе передачи сообщения, а при формировании его отправителем. Во входном сообщении, еще до его передачи по каналу связи, могут иметься искажения (причем не обязательно “ошибки”), обусловленные *неустранимой неоднозначностью* синтаксического представления одного и того же по содержанию сообщения.

Упомянутая неоднозначность может пониматься по-разному. Например, можно трактовать множество всех возможных сообщений как разбитое на некие классы эквивалентности так, что одна

и та же информация может быть представлена разными последовательностями кодов, причем должны существовать алгоритмы выявления эквивалентных кодов. Можно, применяя иную терминологию, считать, что каждое сообщение характеризуется некоторой “размытостью”, а именно содержит, наряду с рядом “существенных” элементов, позволяющих однозначно или с высокой вероятностью отнести его к данному классу, какие-то второстепенные или незначимые элементы.

«Гипотеза неустранимой неоднозначности входных сообщений» принимается, в частности, в случаях автоматической сортировки сообщений, автоматического реферирования, обработки естественно-языковых запросов [4...10]. При этом задачей автоматизированной системы передачи и обработки сообщений становится не только выявление/коррекция ошибок передачи, но и устранение неоднозначности входного сообщения – отнесение его, после получения, к тому или иному классу семантически эквивалентных сообщений.

Методы решения последней задачи существенно зависят от принципов, по которым формируются классы эквивалентности на множестве сообщений. Эти принципы иногда допускают точное формальное выражение. Например, если в сообщении закодирована информация о некоторой математической функции, оно синтаксически представимо кодом *любой* формулы, выражающей эту функцию; и классы эквивалентности содержат коды эквивалентных формул (вроде:  $\sin 2x$ ,  $2\sin x \cos x$  и т. п.). Нередки, однако, ситуации, когда принципы формирования классов эквивалентности сами являются “нечеткими”. В этих случаях выбираются различные формализации “нечеткости”.

Часто классы эквивалентности вводятся на основе понятия “расстояния” между сообщениями: эквивалентными считаются сообщения, синтаксически “близкие” друг к другу. В теории кодирования рассматриваются меры близости, связанные с классами операций, преобразующих сообщения; например, мера близости двух сообщений может определяться числом операций заданного класса, необходимых для “превращения” одного в другое. Подобные меры часто сочетаются с принятием вероятностных гипотез о характере возможных искажений сообщения [1...3]. В принципе, подобные методы могут быть применимы и в случаях принятия гипотезы неустранимой неоднозначности.

В некоторых задачах оказались эффективными другие меры близости сообщений, рассмотренные ниже. Они разработаны исходя из допущений не вероятностного, но иного “качественного” характера возможных искажений сообщения. Этот характер связывается с особой иерархической структурой представления сообщений в интерактивных системах «человек–компьютер».

Представим себе ситуацию компьютерного контроля знаний, когда компьютер ставит вопрос «*Каково наиболее значительное достижение в области математики конца XVII в.; и кто его автор?*», и от тестируемого (студента) ожидается ответ – «эталонное сообщение»:

*Ньютон и Лейбниц изобрели математический анализ.* (\*)

Это сообщение трактуется как модель вывода по аналогии. Допустим, что разрешен ввод ответа на натуральном языке и поступивший от студента ответ (прототип аналогии) таков:

*Матем. анализ – это изобретение и Ляйбница, и Невтона.* (\*\*)

Сравнивая (\*) и (\*\*), мы видим, что сообщение (\*\*) содержит дополнительные слова, сокращения слов, несколько очевидных ошибок и отклоняющиеся от современного написания имен (подчеркнуты). Порядок слов в ней также изменен по сравнению с (\*). Но очевидно, что (\*\*) – абсолютно правильный ответ на заданный вопрос.

Меры аналогичности сообщений, описанные ниже, позволяют отнести (\*) и (\*\*) к одному и тому же классу эквивалентности (и исключить из этого класса “неподходящие” ответы студента – см. примеры в конце статьи).

**2.2. Математические модели аналогии для конечных иерархических систем с линейным порядком на каждом уровне иерархии.** Мы будем рассматривать “сообщение” как произвольный “объект”  $a$ , состоящий из некоторых суб-объектов (элементов). Каждый из этих элементов может, в свою очередь, состоять из суб-объектов низшего уровня и т.д. Примем, что существует некоторое конечное множество  $M_0$  объектов наинизшего (нулевого) уровня, причем на  $M_0$  определено некоторое бинарное рефлексивное “отношение  $\approx$ , которое мы назовем отношением *подобия (сходства)*”. Элементам из  $M_0$  отвечают элементарные, неделимые символы сообщения. Итак, если два объекта –  $a^0$  и  $b^0$  – принадлежат  $M_0$ , то может быть либо  $a^0 \approx b^0$ , т. е.  $a^0$  сходно с  $b^0$ , либо  $\neg(a^0 \approx b^0)$ , т. е.  $a^0$  несходно с  $b^0$  ( $\neg$  – знак отрицания).

Пусть теперь  $a^1$  и  $b^1$  – объекты “1-го уровня”, “построенные из” суб-объектов, принадлежащих  $M_0$ , в одном из следующих двух смыслов:

1) “построен из” означает “является конечным *множеством* вхождений” (в смысле, что  $a^1 = \{a^0, b^0\}$  и  $b^1 = \{a^0, b^0\}$  являются различными объектами 1-го уровня, но при этом  $\{a^0, b^0\} = \{b^0, a^0\}$ );

2) “построен из” означает “является конечной *последовательностью* вхождений” (в смысле, что  $a^1 = \{a^0, b^0\}$  отличается от  $b^1 = \{b^0, a^0\}$ ).

В [31] введены два понятия “подобия объектов 1-го уровня”: 1) (подобие в *широком* смысле): объекты  $a^1$  и  $b^1$  подобны, если число их подобных суб-объектов достаточно велико; 2) (подобие в *узком* смысле):  $a^1$  и  $b^1$  подобны, если длиннейшая подпоследовательность их подобных суб-объектов, сохраняющая порядок следования этих суб-объектов в составе как  $a^1$ , так и  $b^1$ , достаточно длинна.

Определяются численные меры для каждого из двух видов сходства. Именно, если  $a$  и  $b$  – два объекта одного и того же ненулевого уровня, то

$$F_w(a, b) = d_w(a, b) / \max(|a|, |b|);$$

$$F_s(a, b) = d_s(a, b) / \max(|a|, |b|),$$

где  $d_w(a, b)$  – число сходных суб-объектов у  $a$  и  $b$ ;  $d_s(a, b)$  – длина длиннейшей подпоследовательности сходных суб-объектов в  $a$  и  $b$ ;  $|a|$  – длина (число суб-объектов) объекта  $a$ .

Например, пусть  $M_0$  – множество символов латинского алфавита с обычным отношением тождества в качестве  $\approx$ , и пусть  $a = \text{analogy}$ , а  $b = \text{analogia}$ . Очевидно, что при этом  $F_w(a, b) = F_s(a, b) = 6/8 = 0,75$ . Если же переставить в  $b$  два символа, взяв  $b = \text{naalogia}$ , получим снова  $F_w(a, b) = 6/8 = 0,75$ , но  $F_s(a, b) = 5/8 = 0,625$ .

Наконец, выбираются некоторые положительные числа  $L_n$  ( $n = 1, 2, \dots$ ), и постулируется, что два объекта  $a^n$  и  $b^n$  уровня  $n$  “подобны”, если их мера подобия  $F_w(a^n, b^n)$  (либо  $F_s(a^n, b^n)$ ) больше чем  $L_n$ . Имея это новое понятие подобия (сходства) можно применить вышеприведенные определения функций  $F_w$  и  $F_s$  к объектам уровня  $n+1$  и т. д. Имеет место [32]:

**Теорема 1.** Каждая из функций  $1 - F_w$  и  $1 - F_s$  задает метрическое пространство на множестве объектов заданного уровня.

В [31] описана математическая теория подобия конечных символьных последовательностей, в которой формулируются алгоритмы вычисления вышеуказанных мер  $F_w$  и  $F_s$ , и доказывается их сходимость.

**2.3. Меры аналогичности, основанные на “взвешенном” подобии.** Рассмотренные меры  $F_w$  и  $F_s$  могут быть названы симметричными в двух смыслах: 1) для любого  $k \in \{w, s\}$  имеет место  $F_k(a, b) = F_k(b, a)$ ; 2) суб-объекты в составе любого из объектов  $a$  или  $b$  не различаются по своей “важности”, “значимости” для их супер-объекта.

В [32] указанные меры обобщаются. Принимается, что суб-объектам в составе модели  $a$  можно приписывать различные *веса*. Вес любого суб-объекта отражает его “существенность” для  $a$  *сравнительно* с другими суб-объектами из  $a$ . Например, если слово английского языка *analogy* выступает в качестве объекта-модели аналогии, то можно (в зависимости, конечно, от контекста вывода) счесть его начало *anal* более важным, чем окончание; и приписать вес 2 каждой букве в *anal* и вес 1 всем остальным буквам.

При вычислении степени аналогичности “взвешенных” объектов возникают существенно новые, по сравнению с невзвешенным случаем, факторы. Они связаны с тем, что мы можем взвесить все суб-объекты модели  $a$ , но лишены такой возможности по отношению к тем суб-объектам прототипа  $b$ , которые *отсутствуют* в  $a$ . Поэтому для функций, оценивающих сходство  $a$  и  $b$ , вводится параметр “*ожидаемого веса*” элементов  $b$ , не входящих в  $a$ .

В [32] вводятся две асимметричные меры оценки сходства –  $G_w(a, b, L)$  и  $G_s(a, b, L)$  – базирующиеся на приписывании весов суб-объектам модели аналогии  $a$ . Индексы  $w$  и  $s$  обозначают широкий или узкий смысл подобия – так же, как для функций  $F_w$  и  $F_s$ . Параметр  $L$  – это действительное число, отражающее “ожидаемую важность” тех символов прототипа  $b$ , которые отсутствуют в  $a$  (и поэтому не могут быть “прямо” взвешены).

В [32] показано, что меры  $G_w$  и  $G_s$  обладают рядом свойств, которых “естественно ожидать” от функций оценки сходства:

**Теорема 2.** Меры  $G_k(a, b, L)$  ( $k \in \{w, s\}$ ) имеют следующие свойства:

1) если  $L_1 \leq L_2$ , то  $G_k(a, b, L_2) \leq G_k(a, b, L_1)$ ;

2)  $0 \leq G_k(a, b, L) \leq 1$  для любых  $a, b, L$ ;

3) если  $L > 0$  и  $G_w(a, b, L) = 1$ , то прототип  $b$  является перестановкой тех суб-объектов модели  $a$ , которые имеют ненулевые веса;

4) если  $L > 0$  и  $G_s(a, b, L) = 1$ , то прототип  $b$  совпадает с моделью  $a$  с точностью до некоторых суб-объектов  $a$ , имеющих нулевые веса;

5)  $G_s(a, b, L) \leq G_w(a, b, L)$  для любых  $a, b, L$ .

Как оказалось, алгоритмы для вычисления мер взвешенного и невзвешенного подобия значительно отличаются друг от друга. Во взвешенном случае мы не можем, как ранее, оценить узкое (учитывающее порядок) подобие, используя длину длиннейшей подпоследовательности общих (подобных) суб-элементов модели и прототипа. Здесь необходима не самая длинная подпоследовательность, а та, у которой максимален вес. В [32] предложены алгоритмы, вычисляющие степени взвешенного подобия за полиномиальное время.

**2.4. Меры аналогичности, учитывающие “тонкую” структуру объекта.** Выше рассматривались меры оценки аналогии, связанные с 1) “элементным” составом объектов (в частности, для сообщений-текстов – с их лексическим составом); и 2) с учетом (неучетом) *линейного порядка* элементов в составе объекта.

В [33] описаны численные меры “аналогичности” объектов, предполагающие их более “тонкую” структуру, чем просто линейный порядок слов. “Объекты” = “сообщения”, как и выше, суть иерархически организованные системы элементов произвольной природы, для которых значимым является линейный порядок на каждом уровне иерархии. Но теперь он будет *не единственно* значимым отношением на указанном уровне.

Будем считать, что в сообщении по каким-либо основаниям выделяются особые *группы* слов-“имён” (например, в русском языке это могут быть группа подлежащего и группа сказуемого в предложении; в алгоритмическом языке – группы, относящиеся к различным операторам программы, и т. п.). Принимается следующий “*принцип связности*” для групп: *перестановки слов внутри групп “разрушают” структуру сообщения в меньшей степени, чем чередование слов, принадлежащих разным группам.*

Пусть даны два сообщения –  $\mathbf{A}$  и  $\mathbf{B}$ , – где  $\mathbf{B}$  рассматривается как прототип, а  $\mathbf{A}$  – как модель аналогии. Для модели  $\mathbf{A}$  задано множество групп имен  $\{G_i\} i \in [1, n]$ . Говоря неформально, для оценки близости  $\mathbf{A}$  и  $\mathbf{B}$  проверяются следующие основные условия:

1) имеют ли  $\mathbf{A}$  и  $\mathbf{B}$  сходный лексический состав;

2) в какой степени группы имен прототипа  $\mathbf{B}$ , соответствующие группам  $\{G_i\}$  из  $\mathbf{A}$ , “засорены” чужеродными именами.

Аналогично понятию “группы имен в сообщении” можно ввести понятие группы символов в слове (например, группы букв, относящихся к корню слова в этнических языках, и т. п.). Численные меры, оценивающие условие 2, базируются на следующем понятии расстояния между символами  $a$  и  $b$  из группы  $G_i$  в слове  $\mathbf{W}$ .

Мерой “разобщенности” символов  $a$  и  $b$  в слове  $\mathbf{W}$  называется [33] число  $\rho(a, b)$  элементов  $e_k$ , расположенных в  $\mathbf{W}$  между  $a$  и  $b$ , таких что  $e_k \in \mathbf{U} \setminus G_0 \setminus G_i$ . Здесь  $\mathbf{U}$  – некоторое “универсальное” множество символов, а  $G_0$  – особая группа “нейтральных” символов (которые “не влияют на разобщенность”). Имеет место [33].

**Теорема 3.** Мера разобщенности  $\rho(a, b)$  является расстоянием на подмножестве элементов множества  $G_i$ , входящих в слово  $\mathbf{W}$ .

Пусть  $G_i[\mathbf{W}]$  – последовательность всех символов группы  $G_i$  в слове  $\mathbf{W}$ . Пусть длина  $G_i[\mathbf{W}]$  равна  $v \geq 1$ , а  $\mathbf{M}$  – матрица размерности  $v \times v$ , элементы которой – разобщенности  $\rho(a, b)$  всех элементов множества  $G_i[\mathbf{W}]$ .

Средней разобщенностью  $\mu(G_i[\mathbf{W}])$  множества  $G_i[\mathbf{W}]$  называется [33] число 0 при  $v = 1$ , а при  $v > 1$  – число

$$\mu(G_i[\mathbf{W}]) = \frac{\sum_{q=1}^v \sum_{r=1}^v M_{q,r}}{v(v-1)}.$$

Кроме того, в [33] введено понятие *спектра* разобщенности элементов группы  $G_i$  в слове  $\mathbf{W}$ . Спектр – это последовательность натуральных чисел  $\langle L_1, \dots, L_{v-1} \rangle$ , где  $L_k = \rho(a_k, a_{k+1})$ ,  $k \in [1, v-1]$ . Имеет место [33]:

**Теорема 4.**

$$\mu(G_i[\mathbf{W}]) = \frac{2}{v(v-1)} \sum_{j=1}^{v-1} j(v-j)L_j.$$

В [33] получен ряд оценок зависимости меры  $\mu(G_i[W])$  от длины  $\nu$  последовательности  $G_i[W]$ , свойств спектра разобшенности и других параметров.

Приведем примеры применения рассмотренных мер к оценкам аналогичности различных предложений русского языка относительно упоминаемого в разд. 2.1 сообщения (\*).

В “эталонном” сообщении (\*) выделяются именные группы:  $G_1$  – множество слов, подобных одному из слов {*Ньютон, Лейбниц*};  $G_2$  – множество слов, подобных для {*изобрел, открыл, придумал, создал, автор*};  $G_3$  – множество слов, подобных для {*математический, анализ, матанализ*}. При этом задаются численные уровни подобия слов, чтобы, например, слова *Ляйбниц* или *Лейбницем* считались эквивалентными слову *Лейбниц*.

По значениям меры  $\mu(G_i[W])$  нижеследующие тексты – прототипы аналогии – ранжируются так (в скобках указаны степени аналогичности “эталону”):

“Близкие” аналогии:

<i>Лейбниц и Ньютон – изобретатели матанализа.</i>	(1.00)
<i>Мат. анализ был создан Ньютоном и Лейбницем.</i>	(1.00)
<i>Лейбниц, а также Ньютон, открыли анализ.</i>	(0.85)
<i>Мат. анализ Ньютон открыл, и Лейбниц.</i>	(0.80)
<i>Матанализ, Лейбниц, Ньютон.</i>	(0.95)

Тексты с низкой степенью аналогичности:

<i>Математика Ньютона открыла Лейбницев анализ.</i>	(0.39)
<i>Открыт и Ньютон анализом Лейбница.</i>	(0.65)
<i>Ньютон придумал Лейбница и создал анализ.</i>	(0.63)
<i>Ньютон открыл математику, Лейбниц – анализ 'и'.</i>	(0.46)

Мы видим, что: 1) хотя все перечисленные сообщения имеют сходный лексический состав; 2) “близкие” аналогии, в противоположность “далеким”, семантически эквивалентны “эталонному” сообщению (\*). Заметим, что данные оценки аналогичности можно улучшить, если ввести дополнительно группу “нейтральных” имен  $G_0$ , отнеся в нее союзы, слово *также*, и т. п.

Разумеется, упомянутый выше “принцип связности”, как и иные принятые допущения о характере возможных искажений сообщений, не всегда приемлемы (например, когда входные сообщения представляют собой не тексты, а формулы). Для таких случаев разрабатываются иные алгоритмы оценки аналогичности. Компьютерная система тестирования знаний CONTROL использует библиотеку процедур вывода по аналогии, применяемых к ответам различных “узких” типов, что позволяет игнорировать несущественные ошибки и вариации ответов и оценивать ответ, как показали эксперименты [34], с той же степенью адекватности, с какой это делает преподаватель при устном или письменном опросе.

В заключение отметим следующее. Рассмотренные примеры применения описанных методов оценки близости сообщений показывают, что в условиях неустранимой неоднозначности входных сообщений можно использовать выводы по аналогии для учета достаточно тонких различий сообщений и их правильной идентификации.

## Литература

1. Питерсон У., Уэлдон Э. Коды, исправляющие ошибки. – М.: Мир, 1976. – 596 с.
2. Скляр Б. Цифровая связь. – М.: Изд. дом «Вильямс», 2003. – 1104 с.
3. Кудряшов Б. Д. Характеристики и алгоритмы декодирования сверточных кодов в системах связи. Дис. в виде доклада ... докт. технич. наук. – М.: ИППИ РАН, 2004. – 63 с.; <http://guap.ru/dept03/caf35/referat.pdf>
4. Методы и средства автоматизированного проектирования прикладной онтологии / Б.В. Добров, Н.В. Лукашевич, О.А. Невзорова, Б.Е. Федунцов // Известия РАН. Теория и системы управления. – 2004. – № 2. – С. 58-68.
5. Hobbs J.R., Stickel M., Appelt D., Martin P. Interpretation as Abduction // Artificial Intelligence. – 1993. – 63. – No. 1-2. – P. 69-42; <http://citeseer.nj.nec.com/hobbs90interpretation.html>
6. Любарский Ю.Я. Интеллектуальные информационные системы. – М.: Наука, 1990. – 232 с.
7. Гладун В.П. Естественный язык в целенаправленных системах // Диалог'2000. – 2000. – С. 99-102; [http://www.iprinet.kiev.ua/gf/ii\\_sist.htm#o2](http://www.iprinet.kiev.ua/gf/ii_sist.htm#o2)
8. Михаилян А. Некоторые методы автоматического анализа естественного языка, используемые в промышленных продуктах // ИНТЕЛТЕК. Искусственный интеллект. – 2000; <http://www.inteltec.ru/publish/articles/textan/natlang.shtml>
9. Хан У., Мани И. Системы автоматического реферирования // Открытые системы. – 2000. – Вып. 12; <http://www.osp.ru/text/302/178370/>

10. Карпова И.П. Исследование и разработка подсистемы контроля знаний в распределенных автоматизированных обучающих системах: Дис. ... канд. тех. наук. – М.: Моск. гос. ин-т электроники и математики. – 2002. – 200 с.; [orel.rsl.ru/dissert/karpova\\_i\\_p/dis\\_karpova.pdf](http://orel.rsl.ru/dissert/karpova_i_p/dis_karpova.pdf)
11. Уёмов А.И. Логические основы метода моделирования. – М.: Мысль, 1971. – 311 с.
12. Niiniluoto I. Analogy and inductive logic // *Erkenntnis*. – 1981. – Vol.16. – P. 1-34.
13. Булитко В.К. Моделирование процессов функционирования производственных и экономико-экологических систем. – К.: Наукова думка, 1986. – 184 с.
14. Bledsoe W.W. A precondition prover for analogy // *BioSystems*. – 1995. – 34. – P. 225-247; [ftp://ftp.cs.utexas.edu/pub/bshults/ATP-tech-reports/atp-116.ps](http://ftp.cs.utexas.edu/pub/bshults/ATP-tech-reports/atp-116.ps)
15. Bourelly C., D'efourneaux G., Peltier N.. Building proofs or counterexamples by analogy in a resolution framework // *Proceedings of JELIA 96*. – LNAI 1126. – Springer, LNAI. – 1996. – P. 34-49; <http://citeseer.ist.psu.edu/bourelly96building.html>
16. French R.M. The Computational Modeling of Analogy-Making // *Trends in Cognitive Sciences*. – 2002. – 6(5) . – P. 200-205; <http://www.ulg.ac.be/cogsci/rfrench/analogy.tics.pdf>
17. Kokinov B., French, R. M. Computational Models of Analogy- making // *Encyclopedia of Cognitive Science*. Vol. 1. – London: Nature Publishing Group, 2003. – P. 113-118; [www.u-bourgogne.fr/LEAD/people/french/analogy.ECS.kokinov\\_french.pdf](http://www.u-bourgogne.fr/LEAD/people/french/analogy.ECS.kokinov_french.pdf)
18. Gentner D., Kurtz K.J. Relations, Objects, and the Composition of Analogies // *Cognitive Science*. – 2006. – 30. – P. 609-642.
19. Уинстон П. Искусственный интеллект. – М.: Мир, 1980. – 520 с.
20. Люгер Дж. Ф.. Искусственный интеллект: стратегии и методы решения сложных проблем. – М.: Изд. дом "Вильямс", 2003. – 864 с.
21. Sgurev V., Jotsov V. On network flow interpretations of defeasible inference schemes // 2000; <http://inftech.webservis.ru/it/conference/isanditc/2000/section2/eng/areng5.html>
22. Zagoruiko N.G. From knowledge discovery to knowledge mining // *Искусственный интеллект – 2002*. Материалы Междунар. науч. технич. конф. – Таганрог: Изд-во ТРТУ, 2002. – С.95-100.
23. Гасфилд Д. Строки, деревья и последовательности в алгоритмах: Информатика и вычислительная биология. – С.Пб.: Невский Диалект; БХВ-Петербург, 2003. – 654 с.
24. Spiers G.F. An Analogical Reasoning Based Mathematics Tutoring System // 1996; <http://citeseer.nj.nec.com/spiers96analogical.html>
25. Cheng B.H.C., Jeng J.-J. Reusing Analogous Components // *Knowledge and Data Engineering*. – 1997. – Vol. 9. – No. 2. – P. 341-349; <http://citeseer.ist.psu.edu/jeng94reusing.html>
26. Barbosa S., de Souza C. Making More Sense out of Users' Utterances // 1999; [http://peirce.inf.puc-rio.br/cgilua/cgilua.exe/serg/pub/sim/bs\\_ihc99.pdf](http://peirce.inf.puc-rio.br/cgilua/cgilua.exe/serg/pub/sim/bs_ihc99.pdf)
27. Pilkington R., Hartley R., Hintze D.. Learning to Argue and Arguing to Learn: An Interface for Computer-Based Dialogue Games // *Journal of Artificial Intelligence in Education*. – 1992. – Vol. 3(3). – P. 275-295; <http://cbl.leeds.ac.uk/rachel/papers/jaied.htm>
28. Spanoudakis G., Constantopoulos P. Elaborating Analogies from Conceptual Models // *International Journal of Intelligent Systems*. – 1996. – Vol. 11. – No. 11. – P. 917-974; <http://citeseer.ist.psu.edu/spanoudakis96elaborating.html>
29. Melis E., Veloso M.. Analogy in Problem Solving // 1997; <http://citeseer.nj.nec.com/88636.html>
30. Варшавский П.П., Еремеев А.П. Поиск решения на основе структурной аналогии для интеллектуальных систем поддержки принятия решений // *Известия РАН: Теория и системы управления*. – 2005. – № 1. – С. 97 – 109.
31. Леоненко Л.Л., Поддубный Г.В. Теория подобия конечных последовательностей и ее приложения к распознаванию образов // *Автоматика и телемеханика*. – 1996. – № 8. – С. 119-131.
32. Leonenko L. Analogical inferences in computer assisted knowledge testing systems // *Sixth Multi-Conference on Systemics, Cybernetics and Informatics (SCI 2002)*. – Proc. – 2002. – Vol. XVIII. – P.371–376.
33. Леоненко Л. Л. О мерах аналогичности структур текста // *Современная логика: Материалы VIII Общеросс. науч. конф.* – С.Пб: СПбГУ, 2004. – С. 386–389.
34. Баранов В. Ю. Комп'ютерне тестування з інформатики: підсумки педагогічного експерименту в Одеській національній академії зв'язку // *Теорія та методика навчання математики, фізики, інформатики*. – Кривий Ріг: НМетАУ, 2004. – Вип. 4. – Т. 3. – С. 6-12.