

УДК 81'322.2

*Максименко О.И.*

*Московский Государственный Областной Университет*

## ОЦЕНКА ИНТЕЛЛЕКТУАЛЬНЫХ ЛИНГВИСТИЧЕСКИХ СИСТЕМ МЕТОДОМ НЕЧЕТКОЙ ЛОГИКИ

*Maksimenko O.*

*Moscow State Regional University*

### FUZZY LOGIC AS AN ESTIMATION METHOD OF INTELLECTUAL LINGUISTIC SYSTEMS

**Аннотация.** Автором выдвигается предположение, что плохо формализуемые системы, основанные на естественном языке, требуют методов оценки, отличающихся от традиционных математических. В статье приводится ряд способов оценки и характеристики подобных систем, а также описываются преимущества одного из возможных вариантов оценки эффективности интеллектуальных лингвистических систем с использованием нечеткой логики на примере систем машинного перевода и информационно-поисковых систем. Значительный эмпирический материал подтверждает выводы, к которым приходит автор в ходе данного исследования.

**Ключевые слова:** интеллектуальные лингвистические системы, нечеткая логика, экспертные оценки, системы машинного перевода, информационно-поисковые системы.

**Abstract.** The author supposes that poorly formalizable natural language-based systems demand specific estimation methods different from traditional mathematical ones. In the article a number of estimation methods of linguistic systems is resulted, and also advantages of one of the possible variants of an estimation of intellectual linguistic systems efficiency using the method of fuzzy logic are described by the example of machine translation systems and information retrieval systems.

**Key words:** intellectual linguistic systems, fuzzy logic, expert estimation, machine translation, information retrieval systems.

Начало активных отечественных исследований в области вычислительной лингвистики, машинного перевода, систем формализации естественного языка относятся ко второй половине 20-го века, когда стало понятно, что привычным человечеству способом не то, что проанализировать, даже охватить объем непрерывно появляющейся новой информации стало невозможно. В то время еще не шла речь об интеллектуальных лингвистических системах, к машинному переводу относились как к одной из существенно развившихся за годы Второй мировой войны систем дешифровки. Но у истоков создания подобных систем стояли талантливые и амбициозные молодые лингвисты, получившие традиционное филологическое образование, но понимавшие, что новое время диктует новые подходы и решения к возникшим вызовам времени. К плеяде этих ученых относятся Э.И. Королев, Ю.Н. Марчук, Л.Л. Нелюбин, внесшие заметный вклад в развитие прикладной или, как ее еще называли, инженерной лингвистики.

Вопрос автоматизированной обработки информации на естественном языке оставался (и продолжает оставаться) одним из приоритетных в вычислительной и компьютерной лингвистике, а докторская диссертация Л.Л. Нелюбина «Информационно-статистические и инженерно-лингвистические особенности языка и текста в условиях спецкоммуникации», защищенная в 1975 г., стала существенным вкладом в решение этой проблемы. Разработанный

им в те годы Машинный автономный отраслевой словарь (МАОС-1) и его модификации – МАОС-2 и МАОС-3 подробно изложены в книге «Перевод и прикладная лингвистика» [3]. Эти системы были первой в стране автономной системой искусственного интеллекта, которая вошла в историю машинного перевода, опубликованную в США В.Дж. Хатчинсом [4].

За прошедшие с того времени почти 40 лет компьютерные технологии и напрямую связанная с ними прикладная лингвистика шагнули далеко вперед, но идеи, заложенные в ранних работах исследователей, стоявших у истоков этого направления, не устарели.

Известно, что естественный язык относится к плохо формализуемым системам, поэтому при создании интеллектуальных лингвистических систем и оценке их эффективности входные условия можно задать лишь приближенно и, соответственно, ответы также получаются приближенными. Более того, в подавляющем большинстве задач входные условия и критерии оценок непрерывно изменяются, и желательно определить некоторую степень точности решения задачи, соблюдение которой требуется в конкретном случае. Во многих плохо формализуемых требуются аппарат, с помощью которого можно проводить необходимые вычисления, причем довольно быстро и с приемлемой точностью.

Достаточно давно логиками и математиками были предложены весьма разнообразные методы, способные в определенной степени решить эту задачу. В ряду этих методов находятся метод решения «по правилам», применяемый в техническом анализе; метод «по примерам» реализован в нейросетевых технологиях и показывает высокие результаты точности оценок и прогнозов; подход «по моделям» реализован в современной «теории хаоса» и позволяет оценивать события, качественно изменяющиеся за малые промежутки времени. Если правил, примеров и моделей достаточно много, возникают принципы действия объекта – «правила взаимодействия правил (примеров, моделей и т.д.)»,

т.е. можно оценивать и управлять объектом не только на микроуровне (правила), но и на макроуровне (принципы). Такой подход «на принципах» реализован с помощью нечеткой математики в разнообразных инструментальных системах: от несложных электронных таблиц до совершенных экспертных систем. При решении конкретной задачи предполагается комбинация перечисленных выше подходов.

Если точно задать входные параметры задачи невозможно, то на помощь приходит нечеткая логика, в основе которой лежит теория нечетких множеств, изложенная в серии работ Л. Заде в 1965-1973 годах. В этих работах рассматриваются элементы множеств, для которых функция принадлежности представляет собой не жесткий порог (принадлежит/не принадлежит), а плавную сигмоиду (часто упрощаемую ломаной линией), проходящую все значения от нуля до единицы [5]. Нечеткая логика (fuzzy logic) является надмножеством классической булевой логики. Она расширяет возможности классической логики, позволяя применять концепцию неопределенности в логических выводах. Аппарат нечеткой логики столь же строг и точен, как и классический, но вместе со значениями «ложь» и «истина» он позволяет оперировать значениями в промежутке между ними.

У систем, построенных на базе нечеткой логики, есть ряд отличительных особенностей:

1) возможность оперирования входными данными, заданными нечетко, например, непрерывно изменяющимися во времени значениями или значениями, которые невозможно задать однозначно;

2) возможность формализации нечетких критериев сравнения, таких как «большинство», «возможно», «предпочтительно», «хороший» и пр.;

3) возможность проведения качественных оценок как входных данных, так и выводимых результатов, при этом исследователь оперирует не только самими значениями данных, но степенью их достоверности и ее распределением;

4) возможность быстрого моделирования сложных динамических систем и их сравнительного анализа с заданной степенью точности.

Логико-лингвистические методы описания этих систем основаны на том, что поведение исследуемой системы описывается на естественном (или близком к естественному) языке в терминах лингвистических переменных.

Входные и выходные параметры системы рассматриваются как лингвистические переменные, а качественное описание процесса задается совокупностью высказываний следующего вида:

$L_1$ : если  $\langle A_1 \rangle$  то  $\langle B_1 \rangle$ ,  $L_k$ : если  $\langle A_k \rangle$  то  $\langle B_k \rangle$ , где  $\langle A_i \rangle, i=1,2,\dots,k$  – составные нечеткие высказывания, определенные на значениях входных лингвистических переменных, а  $\langle B_i \rangle, i = 1,2,\dots,k$  – высказывания, определенные на значениях выходных лингвистических переменных.

Совокупность импликаций  $\{L_1, L_2, \dots, L_k\}$  отражает функциональную взаимосвязь входных и выходных переменных и является основой построения нечеткого отношения, заданного на произведении универсальных множеств входных и выходных переменных. Если на множестве  $X$  задано нечеткое множество  $A$ , то композиционное правило вывода определяет на  $Y$  нечеткое множество  $B$  с функцией принадлежности. Таким образом, композиционное правило вывода в этом случае задает закон функционирования нечеткой модели системы [1].

Основой нечеткой экспертной системы для оценки какой-либо интеллектуальной системы, использующей естественный язык, является база нечетких правил логического вывода, которые формулируются, исходя из логических соображений. Система обрабатывает все правила вывода одновременно, определяя степень их влияния на конечный результат, каковым является итоговая оценка качества работы таких интеллектуальных систем как, например, системы машинного перевода (СМП) или информационно-поисковые системы (ИПС).

При разработке базы правил определяется критерий оценки системы (входная переменная), т.е. качество работы может быть оценено, в частности, по количеству ошибок, допущенных при переводе того или иного текста или при поиске документов определенной тематики. Другим критерием может служить характер ошибок, допускаемых системой машинного перевода: грубые ошибки, несущественные и т.д. Ошибки при автоматическом синтезе в машинном переводе бывают грамматическими, синтаксическими и семантическими, следовательно, проводить оценку качества работы системы можно отдельно по каждому типу ошибок, т.е. переменные могут меняться. Такой подход претендует на универсальность: нечеткие экспертные системы позволяют оценивать качество работы интеллектуальных систем с единых позиций в зависимости от степени сложности требуемой оценки [2].

Одним из самых простых вариантов оценки качества машинного перевода является вариант, опирающийся на количество ошибок, допущенных при переводе. Количество ошибок определяет эксперт-филолог. Для проведения экспертизы оцениваемой СМП предъявляется заранее подобранный контрольный текст, впоследствии результат работы СМП обрабатывается экспертом, после чего количество ошибок вводится в экспертную систему, которая и «проставляет» итоговую оценку качества работы системы машинного перевода. Используя один и тот же контрольный текст можно сравнивать различные СМП, находящиеся на одной стадии развития (разработка, отладка, коммерческий продукт). Подобную процедуру можно проводить и по характеру качества ошибок при машинном переводе.

Для количественной оценки качества поиска ИПС необходимо использовать одновременно два частных критерия: критерий полноты и критерий точности поиска. Другим вариантом является использование интегральных критериев, которые представляют собой числовые оценки качества работы ИПС. При использовании частных критериев

могут возникать проблемы, когда, например, у одной выдачи критерий полноты больше, чем критерий точности, а у другой, наоборот, критерий точности больше, чем критерий полноты. В этом случае совместное использование критериев приводит к противоречивой ситуации, когда невозможно отдать предпочтение одной из выдач. Решить указанную проблему позволяет нечеткая экспертная система, входными переменными которой являются критерии полноты и точности поиска, а выходной – уровень качества поиска тестируемой ИПС. В зависимости от предпочтений эксперта, составившего базу правил логического вывода системы, всегда будет сделан однозначный вывод о качестве конкретного поиска. Поскольку оба рассматриваемых критерия качества поиска равнозначны и в равной степени должны влиять на конечный результат, функции принадлежности нечетких подмножеств  $A_i$  и  $B_i$  одинаковы.

Для реализации такого универсального метода оценки качества интеллектуальной системы (СМП или ИПС) необходимо определить входные переменные, задать нечеткие подмножества на множестве значений этих переменных и на множестве значений переменной «качество системы», сформулировать набор правил вывода и предоставить нечеткой экспертной системе конкретные значения входных переменных, определен-

ные экспертом при тестировании системы. В силу того, что результаты правил вычисляются отдельно, нечеткая экспертная система может быть эффективно реализована на базе параллельных вычислительных систем.

Основное преимущество использования идей нечеткой логики заключается в том, что они позволяют относительно быстро и без существенных затрат получать как прогнозные, так и реальные оценки плохо формализуемых систем. Для описания столь непростой и плохо формализуемой сущности как естественный язык, создания и оценки эффективности интеллектуальных лингвистических систем на естественном языке использование нечеткой логики открывает новые возможности.

#### ЛИТЕРАТУРА:

1. Корнеев В.В. и др. Базы данных. Интеллектуальная обработка информации. М.: Нолидж, 2000. 348 с.
  2. Максименко О.И. Формальные методы в современной прикладной лингвистике. М.: Народный учитель, МГОУ, 2002. 254 с.
  3. Нелюбин Л.Л. Перевод и прикладная лингвистика. М.: Высшая школа, 1983. 207 с.
  4. Hutchins W.J. Machine Translation: Past, Present, Future. – Chichester: Ellis Horwood Ltd., 1986. 382 p.
  5. Zadeh L. Fuzzy Sets // Information and Control. 1965. Т. 8 (3). P. 338-353.
-