

**МЕТОДЫ И АЛГОРИТМЫ АВТОМАТИЗАЦИИ
ПСИХОЛИНГВИСТИЧЕСКОЙ МОДЕЛИ ВЕРБАЛЬНОГО
СОЗНАНИЯ, ОСНОВАННОЙ НА АССОЦИАТИВНОМ
И КОГНИТИВНОМ ЭКСПЕРИМЕНТАХ¹⁷**

А. В. Сиренко (Москва)

Московский государственный университет печати им.

И. Федорова

alexander.sirenko@gmail.com

Разработка когнитивных моделей является междисциплинарной задачей, относящейся к философии, когнитологии, психолингвистике. Построение подобных моделей с помощью средств вычислительной техники имеет практическое применение в расширении возможностей автоматизированных систем, модернизации методик обучения, создании естественных интерфейсов человеко-машинного взаимодействия.

Модели вербального сознания описывают прежде всего процессы восприятия и порождения текста, языковой способности человека. В основе подхода к моделированию вербального сознания научной школы «Русская языковая личность» [Ведущая научная школа Ю. Н. Караулова Русская языковая личность, 2011] находится идея когнайзера – базы знаний, интегрирующей процессы вербализации знаний и базы данных ассоциативного и когнитивного экспериментов в форме процедуры осознания [Филиппович, 2007, с. 124]. В моделировании работы когнайзера можно выделить два режима: безальтернативный иллюстрирует работу модели для фиксированной вербализации единицы знания, и режим с альтернативой, который моделирует «угадывание» вербализации единицы знания, аналогично языковой игре «кроссворд».

В качестве формального механизма модели выбрана грамматика общего вида, интегрирующая по оригинальной методике результаты ассоциативного и когнитивного экспериментов. Грамматика позволяет производить вывод на основе неполного совпадения левой части правил с предложением, оперируя более 10^4 правил [Сиренко, 2011]. В режиме моделирования с альтернативой формируется список альтернативных вербальных единиц (символов грамматики), соответствующих исходному запросу моделирования, заданному в виде естественно-языковой пропозиции.

Оценка производится через построение на основе модели бинарного классификатора релевантности альтернативы запросу. В терминологии

¹⁷ Работа выполнена при финансовой поддержке гранта РГНФ №12-04-12039в и гранта Президента РФ №НШ-3661.2012.6.

информационного поиска классификатор должен для некоторого запроса $Query_i$ и потенциального ответа $Answer_j$ определить принадлежность $Answer_j$ к классу релевантных ответов $Rel(Query_i)$. В качестве классификатора выбрана модель линейной регрессии, как регрессионная модель с минимальными требованиями к количеству обучающей выборки.

Оценка производится через построение графика чувствительность-специфичность модели:

$$1. Se = \frac{TP}{TP + FN} * 100\% \text{ (чувствительность модели);}$$

$$2. Sp = \frac{TN}{TN + FP} * 100\% \text{ (специфичность модели);}$$

TP – истинно-положительные результаты классификации;

TN – истинно-отрицательные результаты классификации;

FN – ложно-отрицательные результаты классификации;

FP – ложно-положительные результаты классификации.

3. AUC – площадь под кривой ROC-графа, количественный показатель прогностической силы бинарного классификатора. ROC-анализ – построение графика $Se = F(1-Sp)$ [Панклин, 2010].

График позволяет выбрать порог отсечения для перехода от линейной регрессии к бинарному классификатору. После определения порога модель может использоваться как классификатор соответствия языковой единицы единице знания о мире, заданной в естественно-языковой форме.

Функционирование модели, а также качество построенного классификатора, зависят от ряда параметров, устанавливаемых эмпирически:

- Параметров вывода в грамматике: стратегии ограничения вывода, параметров работы автомата применения правил при неполном совпадении, применении фильтра областей;

- Ограничений на исходные данные построения классификатора / запросов: способа представления запроса (описание, метафора и т. д.), длины пропозиции, когнитивной области (быт, литература, война и т. д.);

- Особенности построения классификатора: типа классификатора, состава прогностических величин, объема обучающей выборки, стратегии выбора уровня отсечки.

Данные параметры могут варьироваться исследователем, чем достигается функционирование модели в различных условиях. Таким образом, автоматизированная система предоставляет исследователю инструментальный для проверки гипотез касательно модели.