Далее мы приводим таблицу, в которой представлены основные типы фреймов, активируемых в двух когнитивных системах при структурировании знания о пространстве.

Типы фреймов

Англоязычная когнитивная система	Русскоязычная когнитивная система
Пространственно-дорсальный	неконкретно
Пространственно-вентральный	Пространственно-вентральный
Пространственно-фронтальный	Пространственно-смежный
Пространственно-арриерный	(проксимальный)

СПИСОК ЛИТЕРАТУРЫ

1. Хайруллин В. И. Структурирование информации о пространстве // НТИ. Сер. 2 .— 1997 .— № 11.— С. 30–36.

2. Fedorowicz-Bacz B. Semantic Problems with "Left" and "Right" // Ed. J. Fisiak. Papers and Studies in Contrastive Linguistics.— Poznan: Adam Mickiewicz Univ., 1976.— Vol. 4 .— P. 241–263.

3. Talmy L. How Language Structures Space // Ed. H. Pick, L. Acredolo. Spatial Orientation: Theory, Research, and Application.— Berkley: Plenum Press, 1983. — P. 225–282.

4. The Concise Oxford Dictionary.— Oxford: Clarendon Press, 1978 .— 1368 p.

5. Селиверстова О. Н., Швейцер А. Д. Пространственно-экзистенциальные предложения в русском и английском языках (их соотношение и перевод) // Текст и перевод.— М.: Наука, 1988 .— С. 130–144.

ПРИЛОЖЕНИЕ

СПИСОК ЦИТИРУЕМЫХ ЛИТЕРАТУРНЫХ ИСТОЧНИКОВ

Braine J. Room at the Top.— М.: Foreign Languages Publishing House, 1961 .— 272 p.

Eastmen Christensen General Catalog, 1991 .— 780 p.

Maugham W. S. The Moon and Sixpence.— М.: Progress Publishers, 1969 .— 240 p.

Orwell G. Nineteen Eighty- Four.— Harmondsworth: Penguin Books, Ltd., 1984 .— 286 p.

Puzo M. The Godfather.— London — Sydney: Pan Books Ltd., 1974 .— 448 p.

Salinger J. D. The Catcher in the Rye.— М.: Progress Publishers, 1979 .— 248 p.

Shaw I. Rich Man, Poor Man.— Sevenoaks, Kent: New English Library, 1985 .— 767 p.

Steinbeck J. Of Men and Mice // American Short Novels: The 20th Century.— М.: Raduga, 1987 .— P. 9–86.

Updike J. Rabbit, Run.— Greenwich, Conn.: Fawcett Publications, Inc., 1962 .— 255 p.

Апдайк Дж. Кролик, беги // Апдайк Дж., Рот Ф. Романы.— М.: Физкультура и спорт, 1991 .— С. 15–233.

Брэйн Дж. Путь наверх.— М.: Изд-во иностр. лит-ры, 1960 .— 263 с.

Моэм С. Луна и грош.— М.: Политиздат, 1990 .— 207 с.

Оруэлл Дж. 1984: Роман // Оруэлл Дж. Проза отчаяния и надежды: Роман, сказка, эссе.— Л.: Лениздат, 1990 .— С. 3–248.

Пьюзо М. Крестный отец //Пьюзо М. Крестный отец. Сицилиец.— Баку: Олимп, 1991 .— С. 3–296.

Стейнбек Дж. О мышах и людях // Стейнбек Дж. Избр. произведения.— М.: Правда, 1988 .— С. 179–266.

Сэлинджер Дж. Над пропастью во ржи // Сэлинджер Дж. Над пропастью во ржи: Повести. Рассказы.— М.: Правда, 1991 .— С. 21–184.

Шоу И. Богач, бедняк. Нищий, вор.— М.: Радуга, 1986.— С. 19–440.

Материал поступил в редакцию 12.05.99

УДК 81'322:004.833.3:519.21

Ю. Г. Зеленков

О совместном использовании метода аналогии и методов теории вероятностей при решении задач компьютерной лингвистики

Обосновывается целесообразность совместного применения метода аналогии и методов теории вероятностей к решению задач компьютерной лингвистики. Показано, что это позволяет не только оценивать уровень надежности процедур автоматического анализа и синтеза текстов, но и совершенствовать эти процедуры.

Дж. Ст. Милль в своей монографии "Система логики. Изложение принципов доказательства в связи с методами научного исследования" пишет: "Умозаключение по аналогии можно привести к следующей формуле: две вещи сходны одна с другой в одном или более отношениях; такое-то положение истинно относительно одной из них; следовательно, оно истинно и относительно другой" [1, с. 447]. При этом он подчеркивает связь аналогии с индукцией, которую он определяет как "... процесс, при помощи которого мы заключаем, что то, что истинно относительно нескольких индивидуумов класса, истинно также и относительно всего класса, или что то, что истинно в известное время, будет истинно, при подобных же обстоятельствах, и во всякое время" [там же, 244].

Умозаключение по аналогии он считает по существу индуктивным умозаключением, но не имеющим значения полной индукции, а на с. 450 характеризует его как менее строгую индукцию. Тем не менее, там же он замечает, что в некоторых случаях умозаключение по аналогии может весь-

ма близко подходить по своей силе к настоящей индукции.

Индуктивные умозаключения и, в частности, умозаключения по аналогии весьма сходны с вероятностными суждениями. На это обстоятельство неоднократно обращали внимание многие ученые, например, С. А. Лебедев [2]. Действительно, когда при статистической обработке данных по распределению одной выборки из генеральной совокупности судят о характере распределения всей генеральной совокупности или о характере распределений всех других выборок из генеральной совокупности, то это по существу является умозаключением по аналогии или, иными словами, неполной индукцией. Разница, может быть, состоит лишь в том, что при вероятностных суждениях обычно стремятся определить количественную меру возможных ошибок, например, с помощью доверительных интервалов и доверительных вероятностей или с помощью других критериев согласия статистических гипотез и эмпирических распределений.

Использование в теории вероятностей количественной меры для оценки надежности статистических гипотез — сильная сторона этой теории. Разумно было бы сопровождать такой мерой и умозаключения по аналогии. Это можно делать, если ввести вероятностную меру надежности умозаключений по аналогии: для каждого умозаключения определять вероятность его истинности.

В связи с этим в работе [3] при решении задач компьютерной лингвистики с помощью метода аналогии предлагается придерживаться такой последовательности действий:

1) сформулировать гипотезу о признаках некоторого класса объектов А, который, в свою очередь, может характеризоваться набором других заранее известных признаков;

2) применить процедуру логического вывода по аналогии, в результате которой гипотетические признаки класса объектов А приписываются конкретным объектам, если их известные признаки совпадают (полностью или частично) с известными признаками класса А;

3) оценить эффективность процедуры вывода путем ее многократного применения к различным объектам и определения вероятности получения правильного результата.

Вероятностная оценка надежности процедур логического вывода по методу аналогии имеет важное значение, но еще важнее найти пути повышения этой надежности. Одним из таких путей может быть следующий. Если при выполнении операции, указанной в п. 3, помечать все объекты, относительно которых делаются неправильные выводы, и составить словарь наименований этих объектов (словарь "стоп-объектов"), то тогда, в дальнейшем, при каждом акте вывода можно сверяться с этим словарем и корректировать результаты вывода. При таком подходе вероятность получения правильного вывода повысится.

В работе [3] рассматриваются возможности применения метода аналогии к широкому спектру задач компьютерной лингвистики в области морфологии, синтаксиса и семантики, но мало внимания уделяется вероятностным аспектам этой проблемы. Мы попытаемся восполнить этот пробел на примере морфологического анализа.

Напомним, что идея морфологического анализа по методу аналогии основана на наличии в ряде языков (например, в русском и английском) сильной корреляции между грамматическими характеристиками слов и буквенным составом их концов (не обязательно из-за наличия у этих слов суффиксов и окончаний) [4].

Гипотеза, которая лежит в основе этого подхода, формулируется следующим образом: слова с одинаковыми конечными буквосочетаниями с высокой вероятностью имеют одинаковые словоизменительные и словообразовательные модели и одинаковые наборы грамматической информации (для русского языка — это часть речи, род, число, падеж, лицо и др.). Исходя из такой гипотезы, грамматическую информацию для "новых" слов можно определять по аналогии со словами, включенными в машинный словарь, при условии, что конечные буквосочетания "новых" слов совпадают с конечными буквосочетаниями слов из словаря.

Процедуру морфологического анализа слов на основе применения метода аналогии можно реализовать с помощью словаря словоформ, в котором каждая словоформа сопровождается набором грамматической информации. Словарь инвертируется и сортируется по алфавиту — представляется в виде обратного инвертированного словаря. При этом последняя буква каждой словоформы ставится на первое слева место, за ней следует предпоследняя и т. д.

В процессе морфологического анализа словоформы текста также инвертируются и ищутся в словаре методом дихотомии. Если они там находятся, то грамматическая информация словарных словоформ переносится на текстовые словоформы, если не находятся, то грамматическая информация назначается текстовым словоформам по аналогии со словарными словоформами. В качестве прототипов выбираются те словоформы словаря, концы которых в наибольшей степени совпадают с концами ненайденных словоформ текста (это будут словоформы, стоящие либо "выше", либо "ниже" мест останова процесса дихотомического поиска при его неудачном завершении).

При реализации процедуры морфологического анализа на ЭВМ нет необходимости хранить в памяти весь буквенный состав словоформ, так как на результаты анализа оказывают влияние только конечные буквосочетания. Выполнив операции, описанные в работе [5], можно сократить исходный грамматический словарь словоформ примерно в десять раз, и это не повлияет на точность морфологического анализа всех первоначально включенных в него словоформ, а точность анализа остальных словоформ языка будет достаточно высокой.

Однако эта точность может быть значительно повышена, если в систему морфологического анализа ввести некоторое подобие словаря "стоп-объектов" в виде перечня неправильно анализируемых словоформ, сопровождаемых правильными наборами грамматической информации. Тогда при анализе слов можно сверяться с этим перечнем и корректировать результаты анализа.

На описанных принципах были построены процедуры автоматического морфологического анализа русских и английских текстов [3], которые работают на ПЭВМ типа PENTIUM со скоростью несколько тысяч слов в секунду и с вероятностью правильного анализа более 0,99.

ВЫВОДЫ

При решении задач компьютерной лингвистики с помощью метода аналогии полезно применять методы теории вероятностей не только для предварительной оценки правдоподобности исходных гипотез и построенных на их основе процедур автоматического анализа и синтеза текстов, но и для совершенствования таких процедур. Для этого необходимо в процессе статистических испытаний процедур помечать все объекты, относительно которых делаются неправильные выводы, и составлять словари наименований этих объектов (словари "стоп-объектов"). В дальнейшем целесообразно использовать эти словари для корректировки результатов вывода по аналогии.

СПИСОК ЛИТЕРАТУРЫ

1. Милль Д. С. Система логики. Изложение принципов доказательства в связи с методами научного исследования.— М.: Изд-во "Книжное дело", 1900.

2. Лебедев С. А. Индукция как метод научного познания.— М.: МГУ, 1980.

3. Белоногов Г. Г., Зеленков Ю. Г., Новоселов А. П., Хорошилов Ал-др А., Хорошилов Ал-сей А. Метод аналогии в компьютерной лингвистике // НТИ. Сер. 2 .— 2000 .— № 1.

4. Белоногов Г. Г. Об использовании принципа аналогии при автоматической обработке текстовой информации // Проблемы кибернетики.— 1974 .— № 28.

5. Белоногов Г. Г., Зеленков Ю. Г. Еще раз о принципе аналогии в морфологии // НТИ. Сер. 2. — 1995 .— № 3.

Материал поступил в редакцию 22.11.99

Наши авторы

ШЕЛОВ Сергей Дмитриевич — доктор филологических наук, ведущий научный сотрудник Комитета научной терминологии в области фундаментальных наук РАН, Москва

СТЕПАНЕЦ Ольга Борисовна — аспирантка кафедры романской филологии Российского государственного педагогического университета им. А. И. Герцена, С.-Петербург

Гольдштейн Сергей Львович — профессор, доктор технических наук, зав. кафедрой вычислительной техники Уральского государственного технического университета, академик МАИ, МАНПО, РАЕН, МАИЗ, Екатеринбург

КУДРЯВЦЕВ Александр Григорьевич — старший преподаватель Уральского государственного технического университета

ТКАЧЕНКО Татьяна Яковлевна — кандидат технических наук, доцент Уральского государственного технического университета

ПОДИНОВСКИЙ Владислав Владимирович — профессор, доктор технических наук, зав. кафедрой высшей и прикладной математики Академии труда и социальных отношений, Москва

РАББОТ Жозеф Михайлович — доцент кафедры высшей и прикладной математики Академии труда и социальных отношений

ХАЙРУЛЛИН Владимир Исханович — доктор филологических наук, профессор, зав. кафедрой делового иностранного языка и перевода Башкирского государственного университета, Уфа

ЗЕЛЕНКОВ Юрий Григорьевич — кандидат технических наук, старший научный сотрудник ВИНИТИ, Москва

Редактор Т. Н. Лаппалайнен

Технический редактор Л. В. Кутакова

ЛР № 021074 от 02.09.96

Подписано в печать 14.02.2000 Сдано в набор 22.12.99

Бум. кн.-журн. Ф-т 60×84 1/8 Печать офсетная Гарн. литерат.

Усл. печ. л. 3,73 Уч.-изд. л. 4,70 Тир. 313 экз. Заказ 30229

Адрес редакции: 125315, Москва, ул. Усиевича, 20. Тел. 152-66-71

Производственно-издательский комбинат ВИНИТИ,

140010, г. Люберцы, 10 Московской обл., Октябрьский проспект, 403. Тел. 554-21-86

ISSN 0548-0027. НТИ. Сер. 2. Информ. процессы и системы. 2000. № 2. 1–32.