где $\lim_{|x|\to\infty} \beta(|x|)/|x|^{2/(m-p)} = 0$, *и* $0 \le A < \left\{c(m-p)^2/[2m(m+p)]\right\}^{1/(m-p)}$. *Тогда в любой точке* $y \in \mathbf{R}$ *обобщенное решение задачи Коши (1), (2) из класса* $\mathbf{K}$ *обращается в ноль за конечное время.*

Для других соотношений показателей доказаны аналогичные теоремы.

Показана определенная точность полученных результатов.

Литература

1. Гладков А.Л. О поведении решений некоторых квазилинейных параболических уравнений со степенными нелинейностями // Мат. сборник. 2000. Т.191. № 3. С.25-42.
2. Храмцов О.В. Относительная стабилизация одного нелинейного вырождающегося параболического уравнения // Дифференциальные уравнения. 2001. Т.37. № 12. С.1650-1654.

СТАТИСТИЧЕСКОЕ ПОНЯТИЕ ТЕРМИНА В КОМПЬЮТЕРНЫХ ТЕКСТАХ

А.В. Пушкарёв

При анализе трудов исследователей можно составить оценку современной ситуации. Качеством учебных текстов занимались Мацковский М.С., Микк Я.И., Flesh R., Gunning R., Kincaid J., Spache G., Fry E. и т.д. Использовали частотный анализ для обоснования структуры учебного содержания целей Григорьев С.Г., Гринщкун В.В., Ермаков А.Е., Хмелёв Д.В.

Можно привести большой список учёных, которые за основу взяли понятия компьютерной лингвистики. Стоит показать связь предыдущих понятий в контексте программирования, в котором исходная задача (то есть задача до составления математической модели, модели с учётом или же, наоборот, с игнорированием некоторых свойств и особенностей) имеет довольно абстрактную и свободную форму. Главную трудность и составляет применение компьютерных методов к конкретной математической модели (даже если данная модель составлена без грубых ошибок и/или незначительных погрешностей). Так и в случае лингвистики.

Как многообразие форм, свойств, закономерностей и особенностей языка можно как-то анализировать при помощи «сухой» машины? На самом деле можно. Необходимо всего лишь разобраться хотя бы в каком-то минимальном наборе ключевых понятий, и самое главное, уточнить преследуемые цели.

По критерию поставленных конечных целей особенного внимания заслуживают:

1. Сайт «RCO» (Russian Context Optimizer) с «сайтообразующим» программным средством Ермакова А.Е., который видит приоритетным производить автоматический компьютерный анализ текста, основываясь на понятии *кореферентности*.
2. Программа Хмелёва Д.В. «Лингвоанализатор 3-ε» для анализа текста и определения его в группу текстов по выбранным критериям, работающая при построении относительной энтропии на основе рассмотрения цепей Маркова. На источник [1] ссылался, в том числе, и Хмелёв Д.В. как на чисто информационное определение энтропии через сложность; где сложностью последовательности букв текста А является длина (в двоичном алфавите) минимальной программы, которая выводит А, а энтропия А – это сложность, делённая на длину А в битах. Раскрытие обобщенного понятия энтропии хорошо происходит в [2].
3. Публикации Гринщкуна В.В.
4. Диссертация Максименко О.И «Формальные методы оценки эффективности систем автоматической обработки текста».
5. Частотно-позиционные фильтры.
6. Методика определения авторства, появившаяся в 1998 г.

Но вернёмся к особенности языков мира, состоящей в том, что они различны по сути, начиная от согласования слов между собой, падежей, написания и заканчивая положением (в том числе и возможностью перестановки) слов в предложении.

Учитывая, что понятие лингвистики на самом деле очень интересно и многогранно – психологи, педагоги и филологи определили особое место для данной темы – следует подходить к выбору методов вычленения базовых элементов и средств их обработки и анализа полученных результатов с особым трепетом. Так как возможным является только предположение о преимуществах или недостатках того или иного способа обработки текстов.

Базовым понятием для анализа дидактической оптимальности компьютерного текста в нашей научной работе был выбран **термин**. Под «термином» понимается слово (вместе с вариан-

тами его словоформ), которое статистически неравномерно распределено по тексту в целом. (А как быть с видоизмененными терминами, поддающимися спряжению, склонению по падежам и т.д. – словоформами?). Проблема отождествления различных словоформ при построении частотного словаря может быть решена следующим способом: в каждом слове оставляется не более 5 литер (байтов), причем все гласные, кроме первой, выбрасываются. Этот прием опирается на традицию некоторых древних языков, который вообще обходится при записи без гласных. Подавляющая информативность именно согласных букв давно известна лингвистам. Об этом факте говорят и "нктр псхлг". Если в словах такого выжатого текста оставить еще и по одной гласной, то "чтен и понмн такг текст станв вобщ тривл". Значит, такое сжатие сохраняет смысл фразы и текста в целом.

Замечено также, что согласные образуют своеобразную решетку узлов, вокруг которой при помощи преимущественно гласных и окончаний, образно говоря, "колеблются" конкретные слова и словоформы: «ИнфОрмАтИкА», «нЕфОрмАльнО» (имеют схожий набор согласных букв «НФРМ..»). Учет одной гласной позволяет различить и эти слова: «Инфрм, Нефрм». Понятны ограничения, присущие данному формализованному подходу. Но он в основном позволил различить по смыслу группы слов и отождествить употребления слов, относящихся к одному понятию. Чисто количественные характеристики такого метода образования частотного словаря следует еще исследовать, так как в рассматриваемых текстах доли потерь и шума при построении словаря составляют порядка процента. Это выражается в формальной омонимии, например, при сжатии слов «ТРАНСлятор» и «ТРАНСпарант» и различении программой словоупореблений «МОДуЛьНый» и «МоДуЛь». Метод построения словарей отбрасыванием псевдоокончаний пока не использовался из-за увеличения времени счета.

Именно используя понятие термина, открывается целый спектр всевозможных методов анализа компьютерного текста, в частности, статистических.

Современные пакеты программ достаточно просто обрабатывают введённые данные с использованием различных процедур, а после обработки уже можно выбрать метод, дающий наилучшие результаты.

Классификация статистических методов:

- по количеству анализируемых признаков (одномерные, двумерные, многофакторные);
- по статистическим принципам, лежащим в основе методов (параметрические и непараметрические);
- по зависимости или независимости сопоставленных выборок.

Литература

1. Колмогоров А.Н. Три подхода к определению понятия «количество информации»//Проблемы передачи информации. 1965. Т.1. №1, С. 3-11.
2. Яглом А.М., Яглом И.М. Вероятность и информация. М.: Наука, 1960.
3. Применение статистических методов анализа данных в научно-исследовательской работе: Учебно-методическое пособие/ Авт.-сост.: А.И. Бочкин, Л.П. Колбасич. – Витебск: Издательство УО «ВГУ им. П.М. Машерова», 2006. – 39 с.
4. Пушкарев, А. В. О некоторых распределениях терминов в компьютерном тексте / А.В. Пушкарев // III Машеровские чтения. Математика. Информатика. Философия. Экономика. Юриспруденция: материалы республиканской научно-практической конференции студентов, аспирантов и молодых ученых, Витебск, 24-25 марта 2009 г. – Витебск: УО "ВГУ им. П.М. Машерова", 2009. – С. 78-79. – Библиогр.: с. 79 (2 назв.).

ИМИТАЦИОННОЕ МОДЕЛИРОВАНИЕ СХЕМ КОРРЕКЦИИ РАВНОВЕРОЯТНОСТНЫХ БЛОКОВ

М.А. Чиркин

Область применения стохастических устройств в технике постоянно увеличивается, а их сложность возрастает. В связи с этим возникает необходимость разработки систем автоматизации моделирования и исследования стохастических устройств [1].

Одними из важнейших компонентов стохастических устройств являются равновероятностные блоки и схемы их коррекции, которые предназначены для формирования последовательностей равновероятностных случайных чисел. Они служат:

- для имитации часто встречающихся равномерно распределенных случайных воздействий;
- в качестве входных параметров при формировании потоков случайных величин, подчиняющихся произвольным распределениям.