БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

**Выпускная работа по  
«Основам информационных технологий»**

**Выполнил магистрант**

**Семиков Андрей Андреевич**

**Минск 2010**

Содержание

Реферат на тему «ИТ в радиофизике» 3

Введение 3

Глава 1. История возникновения. 4

Глава 2. Системный анализ. 5

1. Метод опорных векторов. 5

2. Байесовская сеть. 6

Глава 3. Результаты реализации алгоритмов 8

Заключение 9

Список литературы 10

Приложение А. Класс сортировщика. 11

Действующий личный сайт 15

Тестовые вопросы 15

Список литературы 15

Презентация 15

Приложение А 16

# Реферат на тему «ИТ в радиофизике»

## Введение

Информационные технологии (ИТ, от англ. information technology, IT) — широкий класс дисциплин и областей деятельности, относящихся к технологиямуправления и обработки данных, а также создания данных, в том числе, с применением вычислительной техники.

В последнее время под информационными технологиями чаще всего понимают компьютерные технологии. В частности, ИТ имеют дело с использованием компьютеров и программного обеспечения для хранения, преобразования, защиты, обработки, передачи и получения информации. Специалистов по компьютерной технике и программированию часто называют ИТ-специалистами.

Согласно определению, принятому ЮНЕСКО, ИТ — это комплекс взаимосвязанных научных, технологических, инженерных дисциплин, изучающих методы эффективной организации труда людей, занятых обработкой и хранением информации; вычислительную технику и методы организации и взаимодействия с людьми и производственным оборудованием, их практические приложения, а также связанные со всем этим социальные, экономические и культурные проблемы. Сами ИТ требуют сложной подготовки, больших первоначальных затрат и наукоемкой техники. Их внедрение должно начинаться с создания математического обеспечения, формирования информационных потоков в системах подготовки специалистов.

В следующей главе на базе исторических фактов будет дано объяснение, что радиофизика охватывает практически все упомянутые дисциплины и разделы наук.

## Глава 1. История возникновения.

«Радиофизика» — раздел физики, занимающийся изучением общих закономерностей генерации, передачи, приема, регистрации и анализа колебаний и волн различной физической природы и разных частотных диапазонов, а также их применением в фундаментальных и прикладных исследованиях.

Одним из первых естествоиспытателей, заложивших основы радиофизики, должно быть названо имя Майкла Фарадея, самоучки, ставшего профессором Королевского института в 30 лет. В 1831 году он формулирует закон электромагнитной индукции.

1887 год — Генрих Герц создает источник электромагнитных волн —

вибратор Герца, а в 1888 году экспериментально подтверждает теорию Максвелла, регистрируя существование электромагнитных волн и устанавливая их основные свойства — наличие поляризации, законы отражения и преломления, интерференции. Эксперименты проведены с большой изобретательностью. Закон преломления был исследован с использованием призмы из асфальта с ребром в 1 м, а установленную зависимость интенсивности принимаемого излучения от взаимного расположения металлической печи, находящейся в комнате, где проводились опыты, и приемника излучения можно отнести к началам радиолокации.

1 мая 1885 года — на заседании Русского физико-химического общества А.С. Попов демонстрирует прибор для передачи сигналов на расстояние, «как только будет найден источник электромагнитных колебаний, обладающих достаточной энергией».

Работая над созданием систем управления огнем зенитной артиллерии, Клод Шеннон и Норберт Винер закладывают основы кибернетики. Разрабатываются шифровальные и счетно-решающие машины.

В США начинается разработка атомного проекта Манхэттен. Для расчетов, необходимых при создании атомной бомбы, создается электронная вычислительная машина — «ЭНИАК».

1954 год — Н. Г. Басов, А. М. Прохоров создают и исследуют мазер и в 1964 году вместе с Ч. Таунсом получают Нобелевскую премию за создание мазера.

Человечество вошло в новую эру информационных и коммуникационных технологий, используя достижения микроэлектроники, нелинейной и интегральной оптики, криоэлектроники, перенося идеи радиофизики в микроволновый и оптический диапазоны.

Все науки, специализации, охватывающие работу с радиоволнами, магнитрым полем, можно включить в понятие «радиофизика».

## Глава 2. Системный анализ.

Для решения некоторых задач требуется применение совсем не стандартных методов моделирования. Одним из таких является моделирование взаимодействия нейронов биологического мозга.

Системный анализ как средство установления структурных связей между элементами исследуемой сети нейронов является одним из инструметов радиофизики.

Одной из моих задач в частности был анализ и проектирование искусственной нейронной сети. Анализ заключался в реализации нескольких методов моделирования самообучающейся нейросети сети для сортировки текстовых e-mail сообщений.

### 1. Метод опорных векторов.

Первым решением был метод опорных векторов, который заключался в разделении классов данных с помощью многомерной плоскости.

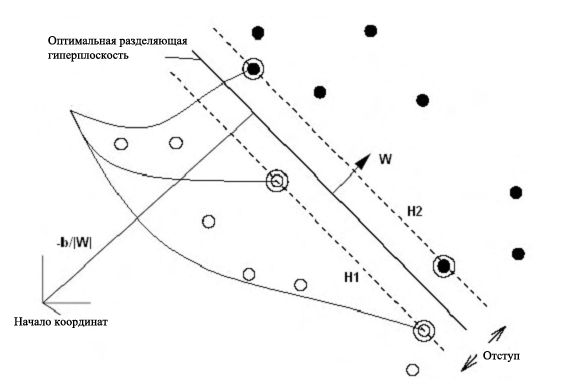


Рисунок 1. Разделяющая гиперплоскость.

Вектором (рис. 1) является перпендикуляр к разделяющей плоскости, называемый также опорным.

Преимуществами данного метода:

•Вместо многоэкстремальной задачи решается задача квадратичного программирования, имеющая единственное решение. Методы оптимизации в этом случае существенно более эффективны.

•Автоматически определяется число нейронов скрытого слоя. Оно равно числу опорных векторов (при использовании ядер).

•Принцип оптимальной разделяющей гиперплоскости приводит к максимизации ширины разделяющей полосы между классами, следовательно, к более уверенной классификации. Градиентные нейросетевые методы выбирают положение разделяющей гиперплоскости произвольным образом, «как придётся».

Недостатки:

•Метод опорных векторов неустойчив по отношению к шуму в исходных данных. Если обучающая выборка содержат шумовые выбросы, они будут существенным образом учтены при построении разделяющей гиперплоскости.

•Дает не более 97% успешного распознания сообщений. Приемлимым считается не менее 99%.

### 2. Байесовская сеть.

Алгоритм фильтра был реализован с использованием web-технологий программирования, таких как фреймворк Symfony 1.4, язык PHP и база данных PostgreSQL 8.4.4 на операционной системе Ubuntu 10.04 Linux.

Программа состоит из основных трех логических блоков:

1. Тренер нейросети
2. Класс разбиения текста на элементарные части
3. Классификатор.

«Наивный» или упрощенный алгоритм Байеса (рисунок 2) является частным случаем байесовской сети, где каждая входная переменная зависит только от переменной класса, которая является единственным корнем графа.

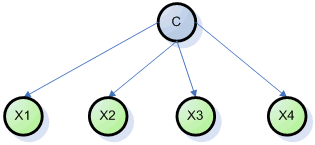


Рисунок 2. «Наивный» или упрощенный алгоритм Байеса.

В Приложении А представлена часть кода программной реализации алгоритма. В данном случае программа классифицирует спам/неспам сообщения электропочты.

## Глава 3. Результаты реализации алгоритмов

Эксперимент с данными на байесовской сети привел к удовлетворительным результатам. Наиболее часто встречаемые слова и предложения были отнесены к спаму с вероятностью более 99%.

Метод опорных векторов дал удовлетворительный результат, однако из-за недостатков, связанных с вычислительными затратами, предпочтение отдано нейросетевому байесовскому классификатору.

Нейросетевой классификатор эффективен ввиду следующих преимуществ:

* достаточно простая схема обучения
* быстрое обучение сети
* простое и такое же быстрое переобучение по сравнению с методом опорных векторов
* относительно небольшие вычислительные затраты
* возможность иметь гибкую систему классификации сообщений. На практике реализована в виде древовидного алгоритма Байеса.
* на сегодняшний день является самым распространенным алгоритмом сортировки и классификации не только текстовых почтовых сообщений, но и других различных объектов, в том числе и графических.

Данная программа может быть применена в широком спектре задач: от фильтрации e-mail сообщений до сортировки текста по классам.

## Заключение

Радиофизика, в частности системный анализ явились основой в области информационных техногогий. В силу междисциплинарности данного раздела физики направления исследований во многом пересекаются с направлениями исследований в других разделах.

В работе по анализу двух алгоритмов сортировки текстовых сообщений был затронут статистический подраздел радиофизики, затрагивающий системный анализ живых биологических тканей. В данной работе моделирования нейронов головного мозга применялись такие информационные технологии как база данных, web-сервер.

На примере реализации двух методов сортировки было показано, как информационные технологии применяются в радиофизике.

## Список литературы

1. Воронцов К.В. Лекции по методу опорных векторов [Электронный ресурс]. – www.ccas.ru/voron/download/SVM.pdf.

2. Mozina M, Demsar J, Kattan M, & Zupan B. Nomograms for Visualization of Naive Bayesian Classifier // In Proc. of PKDD. – 2004. – С.337—348.

* + 1. Гончаров М. Модифицированный древовидный алгоритм Байеса для решения задач классификации: [Электронный ресурс]. – http://www.spellabs.ru/download/AugmentedNaiveBayes.pdf.
    2. Лекции курса общей физики МГУ. [Электронный ресурс]. Уровень доступа: публичный http://www.osc.phys.msu.ru/mediawiki/upload/d/d6/Lecture\_1-RF-historic\_intro\_04-03-09.pdf

## Приложение А. Класс сортировщика.

class spam {

    protected $\_source;

    /\*\*

\* Constructor

\*

\* Because the Spam Class do not know the source, you must

\* define a function that return an array of token and values.

\*

\* @param string $Callback Function name

\*/

    public function \_\_construct($callback='') {

    }

    /\*\*

\* Returns the the posibility to text belogs to "spams"

\*

\* @param $text Text to analize

\* @access public

\* @return true

\*/

    public function isItSpam($text,$type) {

        $ngram = new ngram;

        $ngram->setText($text);

        for($i=3; $i <= 5;$i++) {

            $ngram->setLength($i);

            $ngram->extract();

        }

        $fnc = $this->\_source;

        $ngrams = $ngram->getnGrams();

        $knowledge = $fnc( $ngrams,$type );

        $total=0;

        $acc=0;

        foreach($ngrams as $k => $v) {

            if ( isset($knowledge[$k]) ) {

                $acc += $knowledge[$k] \* $v;

                $total++;

            }

        }

        $percent = ($acc/$total);

        $percent = $percent > 1.0 ? 1.0 : $percent;

        return $percent \* 100;

    }

    public function isItSpam\_v2($text,$type) {

        $ngram = new ngram;

        $ngram->setText($text);

        for($i=3; $i <= 5;$i++) {

            $ngram->setLength($i);

            $ngram->extract();

        }

        $ngrams = $ngram->getnGrams();

        $knowledge = $this->getNgramsFromDB( $ngrams,$type );

        $total=0;

        $acc=0;

        /\*\*

\* N = total number of n-grams used.

\* K = product of all n-grams (values are extracted from knowledge base)

\*

\* H = chi2Q( -2N K, 2N);

\* S = chi2Q( -2N ( (1.0 - ngram(1)) ( 1.0 - ngram(2)) .. ( 1.0 - ngram(N)) ), 2N)

\* I = ( 1 + H - S ) / 2

\*

\*/

        $N = 0;

        $H = $S = 1;

// var\_dump($ngrams); exit;

        foreach($ngrams as $k => $v) {

         if (($index = ngram::staticCheckIfValueExists($knowledge, $v['ngram'])) === false) continue;

// if ( !isset($knowledge[$k]) ) continue;

            $N++;

            $value = $knowledge[$index]['percent'] \* $v['weight'];

            $H \*= $value;

            $S \*= (float)( 1 - ( ($value>=1) ? 0.99 : $value) );

        }

        $H = $this->chi2Q( -2 \* log( $N \* $H), 2 \* $N);

        $S = (float)$this->chi2Q( -2 \* log( $N \* $S), 2 \* $N);

        $percent = (( 1 + $H - $S ) / 2) \* 100;

        return is\_finite($percent) ? $percent : 100;

    }

    /\*\*

\* get ngrams

\*

\* This is function is called by the classifier class, and it must

\* return all the n-grams.

\*

\* @param Array $ngrams N-grams.

\* @param String $type Type of set to compare

\*/

    public function getNgramsFromDB($ngrams, $type){

// var\_dump($ngrams); exit;

     $info = array();

     foreach ($ngrams as $ngram){

     $info[] = $ngram['ngram'];

     }

$q = Doctrine::getTable('KnowledgeBase')->createQuery('kb')->where('kb.belongs = ?', $type)

->andWhereIn('kb.ngram', $info);

foreach ($q->fetchArray() as $item){

$t[] = array('ngram' => $item['ngram'], 'percent' => $item['percent']);

}

$q->free();

return $t;

    }

    public function chi2Q( $x, $v) {

        $m = (double)$x / 2.0;

        $s = exp(-$m);

        $t = $s;

        for($i=1; $i < ($v/2);$i++) {

            $t \*= $m/$i;

            $s += $t;

        }

        return ( $s < 1.0) ? $s : 1.0;

    }

}

# Действующий личный сайт

http://lifeline.su/mag

# Тестовые вопросы

Какие из перечисленных web-броузеров работают на базе HTML рендер машины Webkit?

1. Internet Explorer
2. Google Chrome
3. Netscape
4. Safari
5. Mozilla Firefox
6. Opera

# Список литературы

Список идентичен списку литературы реферата.

# Презентация

Презентация магистерской работы указана в приложении А.

# Приложение А

|  |  |
| --- | --- |
|  |  |
|  |  |
|  |  |

|  |  |
| --- | --- |
|  |  |
|  |  |
|  |  |

|  |  |
| --- | --- |
|  |  |
|  |  |
|  |  |