

УДК 004.021:614.84

ПРИМЕНЕНИЕ АЛГОРИТМОВ НЕЧЕТКОГО ПОИСКА В СИСТЕМАХ МОНИТОРИНГА ЛЕСОПОЖАРНОЙ ОБСТАНОВКИ

М.А. Сонькин, Ю.В. Лещик

Томский политехнический университет

E-mail: catlen@ya.ru

Рассмотрены алгоритмы нечеткого поиска в реляционных базах данных. Показано, что использование стандартных алгоритмов текстового поиска и нечетких запросов применительно к системам мониторинга лесопожарной обстановки является недостаточным эффективным. Предложен подход к реализации нечеткого поиска в реляционной базе данных на основе комбинирования стандартных алгоритмов.

Ключевые слова:

Нечеткий поиск, текстовый поиск, нечеткие запросы, нечеткие множества.

Key words:

Fuzzy search, text search, fuzzy queries, fuzzy sets.

Наиболее часто встречающимся в программировании действием является поиск [1]. Методы поиска могут быть классифицированы различными способами [2]. В большинстве случаев под поиском понимают задачу выборки, в которой требуется найти элемент, удовлетворяющий некоторым условиям [3]. Однако возникают ситуации, когда поиск представляет собой нахождение объектов, похожих, в определенном смысле, на заданный объект. В таких случаях принято говорить об использовании алгоритмов нечеткого поиска.

Во многом возникновение задач нечеткого поиска обусловлено проблемами управления качеством данных, в том числе связанными с наличием как в запросах, так и непосредственно в базах данных, ошибок ввода информации [4]. Рассмотрим случай необходимого использования алгоритмов нечеткого поиска в системах мониторинга лесопожарной обстановки.

Системы мониторинга лесопожарной обстановки представляют собой распределенные системы по учету информации о лесных пожарах. Данные из различных подразделений авиационной и наземной охраны лесов передаются в центральный диспетчерский пункт. В «центральной» системе реализован wcf-сервис, позволяющий принимать данные о горимости подразделений в формате xml. Выбор формата обусловлен наличием различных источников отправки данных: наряду с системами мониторинга лесопожарной обстановки могут быть использованы специализированный веб-интерфейс, а также «ручная» отправка – формирование отчетного документа оператором.

Для xml-файла вводятся специальные символьные обозначения: N, N – дробное число; N – целое число; D – дата (без времени); DT – дата (со временем); S – строка (всегда без кавычек); ID – char (10).

Передаваемый документ может иметь следующий вид:

```
<doc name = "S" version = "1">
<row type = "fire">
```

```
<id>ID</id>
<cOKPO>S</cOKPO>
<dDate>DT</dDate>
<nNumber>N</nNumber>
<pri_kpri id = «ID»>N</pri_kpri>
<ter id=»ID»>N</ter>
<nPl_Obnar>N</nPl_Obnar>
<name_lesh>S</name_lesh>
...
</row>
</doc>
```

Xml-документ состоит из тегов, соответствующих названиям атрибутов таблиц в базе данных. Для организации процедуры поиска необходимы только значения ключевых полей, а именно значения справочников (субъект РФ, код лесопользователя, зона мониторинга и т. д.), а также значения типа «дата/время» или «дата» и имя таблицы (атрибут name тега doc).

Значение ключевых полей передается в следующем виде:

```
<ter id="">69</ter>.
```

Т. е. для элемента ter может быть указано либо значение узла (на примере 69), либо значение атрибута id. Данное правило обусловлено тем, что в таблицах, содержащих справочники, хранятся не только идентификаторы id, но и специализированные коды. Такая схема дает возможность найти, например, субъект Российской Федерации в таблице Tег двумя способами: по уникальному идентификатору записи id, а также по совокупности полей code и nYear. В данном случае code будет содержать в себе значение, заключенное между тегами «ter», а nYear – год, полученный из ключевого поля типа «дата / время» или «дата».

Значения типа «дата/время» должны передаваться в виде:

```
<dDate>15.06.2012 13:25:55< dDate />.
```

После получения xml-файла происходит его проверка на соответствие ряду логических и синтаксических правил, затем осуществляется разбор элементов и поиск объектов в таблицах базы дан-

ных по ключевым полям. В случае если поиск успешно завершён, характеристики найденного объекта заменяются данными из xml-файла. В противном случае пользователю выдается сообщение о том, что искомым объектом не найден. В данной ситуации следует учесть, что ошибки могут содержаться непосредственно в значениях узлов и атрибутов. А поскольку за процессом разбора сообщений следит оператор, необходимо предоставить ему возможность заменить значения характеристик одного из объектов базы данных значениями из xml-файла, т. е. сформировать список «похожих» объектов. Таким образом, возникает задача нечеткого поиска объектов в базе данных.

Существуют различные подходы к реализации нечеткого поиска в реляционных базах данных. Одним из подходов является использование нечеткого текстового поиска: искомым объектом и запись таблицы базы данных преобразуются в строки путем слияния атрибутов, затем с помощью алгоритмов нечеткого поиска строки вычисляются степень их схожести. В данном подходе под поиском по сходству подразумевается отыскание всех слов, для которых расстояние (метрика) до поискового шаблона не превышает заданную величину.

Наиболее популярная из существующих метрик – это функция Левенштейна–Дамерау. Расстояние Левенштейна равно минимальному числу элементарных операций редактирования, необходимых для преобразования одной строки в другую, в том числе операции замены, вставки и удаления одного символа. В модификации расстояния редактирования, предложенной Дамерау, в множество элементарных операций включены транспозиции символов.

На базе метрики Левенштейна–Дамерау построено большое число поисковых алгоритмов, в т. ч.: сигнатурные алгоритмы, алгоритмы n-граммной индексации, строковые trie-деревья, алгоритмы расширения выборки и последовательного перебора [5]. Одним из наиболее эффективных алгоритмов из приведенных является метод N-грамм.

Вторым подходом к реализации нечеткого поиска в реляционных базах данных является использование нечетких запросов. Механизмы нечетких запросов к реляционным базам данных базируются на теории нечетких множеств, которая была впервые предложена Лофти Заде. Характеристикой нечеткого множества выступает функция принадлежности. Обозначим через $\mu(x)$ степень принадлежности элемента x к нечеткому множеству, представляющую собой обобщение понятия характеристической функции обычного множества. Тогда нечетким множеством S называется множество упорядоченных пар вида $S = \{\mu(x)/x\}$, при этом $\mu(x)$ может принимать любые значения в интервале $[0, 1]$. Значение $\mu(x) = 0$ означает отсутствие принадлежности к множеству, 1 – полную принадлежность.

Для переменных, относящихся к непрерывному виду данных, функцию принадлежности удобнее

задать аналитической формулой и для наглядности изобразить графически. Существует свыше десятка типовых форм кривых для задания функций принадлежности. Рассмотрим самые популярные кусочно-линейные на рис. 1: треугольную и трапециевидальную.

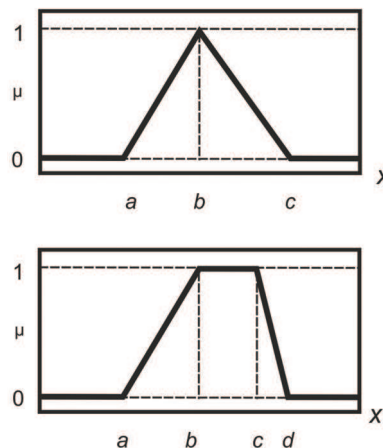


Рисунок. Типовые функции принадлежности

Треугольная функция принадлежности определяется тройкой чисел (a, b, c) , и ее значение в точке x вычисляется согласно выражению:

$$\mu(x) = \begin{cases} \frac{x-a}{b-a}, & a \leq x \leq b, \\ \frac{c-x}{c-b}, & b \leq x \leq c, \\ 0, & \text{иначе} \end{cases} \quad (1)$$

Аналогично для задания трапециевидальной функции принадлежности, представленной формулой (2), необходима четверка чисел (a, b, c, d) :

$$\mu(x) = \begin{cases} \frac{x-a}{b-a}, & a \leq x \leq b, \\ 1, & b \leq x \leq c, \\ \frac{d-x}{d-c}, & c \leq x \leq d, \\ 0, & \text{иначе} \end{cases} \quad (2)$$

Рассмотрим указанные подходы применительно к задаче реализации нечеткого поиска для описанных систем мониторинга лесопожарной обстановки. Для создания алгоритма нечеткого текстового поиска будем использовать метод N-грамм. Словарная N-граммная индексация основана на следующем свойстве: если слово u получается из слова w в результате не более чем k элементарных операций редактирования (за исключением перестановок символов), то при любом представлении u в виде конкатенации из $k+1$ -й строки, одна из строк такого представления будет точной подстрокой w . Это свойство можно усилить, заметив, что среди подстрок представления существует такая, что разность между её позицией в строках w и u

не больше k . Таким образом, задача поиска сводится к задаче выборки всех слов, содержащих заданную подстроку.

Функция нечёткого сравнения использует в качестве аргументов две строки и параметр сравнения – максимальную длину сравниваемых подстрок. Подстроки содержат буквы кириллического алфавита и пробел. Результатом работы функции является число, лежащее в пределах от 0 до 1, где 0 соответствует полному несовпадению двух строк, а 1 – полной их идентичности. Сравнение строк происходит по следующей схеме: функция сравнения составляет все возможные комбинации подстрок с длиной вплоть до указанной и подсчитывает их совпадения. Количество совпадений, разделённое на число вариантов, объявляется коэффициентом схожести строк для фиксированного N и выдаётся в качестве результата работы функции, далее берётся среднее значение для всех коэффициентов. Формула релевантности (3) будет выглядеть следующим образом:

$$R = \frac{\sum_{i=1}^N r(i)}{N},$$

$$r(i) = \frac{\text{Match}(Str1, Str2, i) + \text{Match}(Str2, Str1, i)}{\text{Count}(Str1, i) + \text{Count}(Str2, i)}, \quad (3)$$

где $\text{Count}(Str, i) = (\text{len}(Str) - i + 1)$; $\text{len}(S)$ – длина строки S ; $\text{Match}(S1, S2, i)$ = сумма совпадений всех подстрок длиной i из $S1$ в строке $S2$.

Увеличение длины максимальной подстроки приводит к увеличению времени работы функции. С другой стороны, поиск становится более чётким [6].

Если рассмотреть метод N -грамм применительно к строкам, полученным слиянием атрибутов записей таблиц базы данных и объекта xml, то можно сделать вывод о неэффективности такого подхода. Строковые значения специфичны и представляют собой закодированное сообщение, в них могут совпадать преимущественно 2–4 символа, а для числовых значений гораздо результативнее реализовать поиск с помощью нечетких запросов.

Второй подход реализуем с помощью применения нечеткого отношения ОКОЛО к каждому числовому полю. Т. е. следует задать интервал для определения функции принадлежности по формуле (1) [7].

Например, необходимо вычислить функцию принадлежности кода субъекта 30 ОКОЛО значения кода субъекта 50. Зададим интервал для вычисления функции принадлежности: [20; 50].

Используя треугольную функцию, определим функцию принадлежности:

$$\mu(x) = \frac{x - a}{b - a} = \frac{30 - 20}{80 - 20} = 0,6.$$

Данный подход гораздо эффективнее по отношению к числовым полям, чем алгоритм текстового поиска, однако он неприменим к полям строкового типа.

Таким образом, на основании выявленных особенностей применения приведенных подходов к поиску в системах мониторинга лесопожарной обстановки, предлагается использовать комбинированный алгоритм нечеткого поиска. Поскольку оба вышеизложенных подхода позволяют осуществлять поиск с использованием функций «похожести», необходимо использовать нечеткий текстовый поиск для нахождения строковых полей, а механизм нечетких запросов – для нахождения числовых полей и полей типа «дата/время».

В общем виде алгоритм нечеткого поиска записей в таблице базы данных можно представить в виде следующих шагов:

1. Выделяется первая запись в базе данных.
2. Рассматривается первый поисковый признак, сформулированный в запросе, и его значение.
3. По признаку находится соответствующий атрибут базы данных.
4. Определяется тип атрибута: если атрибут строка, то осуществляется переход на шаг 5, если атрибут является значением даты и/или времени, то – переход на шаг 6, в случае числового значения атрибута – переход на шаг 7.
5. Искомое строковое значение из xml, т. е. атрибут id, состоящий из 10 символов, разбивается на 5 N -грамм по 2 символа. Далее для значений соответствующего строкового атрибута в таблице базы данных вычисляется функция релевантности по формуле (3). Значение функции релевантности сохраняется. Переход на шаг 8.
6. Значение дата и/или время преобразуется в число, затем происходит переход на шаг 7.
7. Вычисляется значение функции принадлежности, соответствующей значению признака, в точке, представляющей значение анализируемого атрибута. Полученное значение функции принадлежности запоминается.
8. Рассматривается следующий поисковый признак, и повторяются шаги 3 и 4 до окончания поисковых признаков. В результате будет получен набор значений функций принадлежности и релевантности всех атрибутов анализируемой записи.
9. Основываясь на полученных наборах значений, вычисляется обобщенная оценка принадлежности анализируемой записи объекту из xml (минимальное значение из указанного набора, произведение элементов набора и т. п.). Полученное обобщенное значение запоминается в рабочем поле базы данных.
10. Осуществляется переход на следующую запись и повторяется шаг 2. Повторение происходит до тех пор, пока не будут перебраны все записи. Результатом поиска по нечеткому запросу станет упорядоченная выборка записей по степени их соответствия данному запросу от 1 (полное соответствие) до 0 (полное несоответствие) [8. С. 64–65].

Выводы

Для систем мониторинга лесопожарной обстановки предложен алгоритм с учетом особенностей

типов данных, основанный на комбинировании базовых механизмов текстового поиска методом N -грамм и нечетких запросов. Показано, что алгоритмы нечеткого текстового поиска и нечетких запросов для нахождения записей в реляционных базах данных систем мониторинга не эффективны, поскольку таблицы содержат поля различных типов, в том числе строковые и численные типы, а также

значения «дата/время». Применительно к строковым значениям наиболее результативным является алгоритм нечеткого текстового поиска на базе метрики Левенштейна–Дамерау, поиск численных значений, в которые можно преобразовать поля «дата-время», целесообразно осуществлять с помощью нечетких запросов, основанных на теории нечетких множеств Лотфи Заде.

СПИСОК ЛИТЕРАТУРЫ

1. Вирт Н. Алгоритмы и структуры данных. – М.: Мир, 1989. – 360 с.
2. Кнут Д.Э. Искусство программирования. Сортировка и поиск. 2-е изд. – М.: Вильямс, 2007. – Т. 3. – 824 с.
3. Макконел Дж. Основы современных алгоритмов. 2-е изд. – М.: Техносфера, 2004. – 368 с.
4. Караханов Д.С. Использование алгоритмов нечеткого поиска при решении задач обработки массивов данных в интересах кредитных организаций // Аудит и финансовый анализ. – 2010. – № 2. URL: www.auditfin.com/2010/2/toc.asp (дата обращения: 15.05.2012).
5. Бойцов Л.М. Классификация и экспериментальное исследование современных алгоритмов нечеткого словарного поиска // Электронные библиотеки: перспективные методы и технологии, электронные коллекции: Труды VI Всеросс. научн. конф. (RCDL'2004). – Пушкино, Россия, 2004. URL: <http://rcdl.ru/doc/2004/paper27.pdf> (дата обращения: 20.05.2012).
6. Караханов Д.С. Использование алгоритмов нечеткого поиска при решении задачи устранения дубликатов в массивах данных // Молодой ученый. – 2010. – Т. 1. – № 8 (19). – С. 150–155.
7. Потапов Е.Н. Нечеткие множества в хранилище данных. 2011. URL: <http://разработка-хд.рф/blog/?p=346> (дата обращения: 01.05.2012).
8. Рыжов А.П. Модели поиска информации в нечеткой среде. – М.: Изд-во ЦПИ при ММФ МГУ, 2004. – 96 с.

Поступила 14.09.2012 г.

УДК 621.313.333:62–83

ИДЕНТИФИКАЦИЯ ПАРАМЕТРОВ АСИНХРОННЫХ ЭЛЕКТРОДВИГАТЕЛЕЙ С НЕПОДВИЖНЫМ КОРОТКОЗАМКНУТЫМ РОТОРОМ НА ОСНОВЕ РАЗНОСТНЫХ СХЕМ

А.С. Глазырин, Е.В. Боловин

Томский политехнический университет
E-mail: asglazyrin@tpu.ru

Разработан метод идентификации параметров динамических моделей асинхронных электродвигателей с неподвижным короткозамкнутым ротором на основе решения системы разностных уравнений. Для цифрового дифференцирования применялась многоточечная аппроксимация первых и вторых производных. Были определены с допустимой погрешностью параметры модели асинхронного двигателя. Доказана работоспособность и эффективность динамической идентификации при использовании разработанного метода.

Ключевые слова:

Асинхронный двигатель, идентификация параметров, разностные схемы.

Key words:

Induction motor, parameters identification, difference schemes.

Введение

Современное производство не может обойтись без электропривода. В электроприводах применяются различные двигатели как по структуре и назначению, так и по возрастной категории. Эффективное управление данными двигателями возможно лишь, если известны текущие значения их параметров. Но зачастую определить параметры либо невозможно, либо крайне сложно. Соответственно, остро встает вопрос об идентификации параметров электродвигателей.

Под идентификацией модели в общем случае понимают определение структуры и ее параметров путем анализа входных и выходных сигналов, по-

даваемых на модель объекта. Основное требование, предъявляемое к процедуре динамической идентификации параметров, является получение несмещённых оценок, стремящихся к истинным значениям искомых параметров. Исходя из этого требования, значительный интерес представляют методы, основанные на решении разностных уравнений, описывающих динамику электродвигателей, или так называемые разностные схемы.

Цель представленной работы – разработать метод идентификации параметров асинхронного двигателя (АД) с неподвижным короткозамкнутым ротором на основе разностных схем.