

УДК 004.04

Н.Г. Марков, Д.М. Сонькин, Т.Т. Газизов, Ю.В. Лещик, А.С. Фадеев, А.О. Шемяков

Комбинированный алгоритм прогнозирования дорожной обстановки на основе методов нечеткого поиска в региональной навигационно-информационной системе мониторинга и управления транспортом

С развитием вычислительной техники и ростом ее производительности подготовка управленческих решений в автоматизированных системах становится отдельным классом задач. Применительно к системам мониторинга и управления транспортом можно выделить множество задач, автоматизированное решение которых позволяет существенно повысить эффективность работы диспетчерской службы. В рамках данной статьи рассматривается задача прогнозирования развития дорожной обстановки (прогнозирование заторов). На основе известных подходов нечеткого поиска в реляционных базах данных предложен комбинированный алгоритм прогнозирования развития дорожной обстановки.

Ключевые слова: нечеткие множества, алгоритм нечеткого поиска, автоматизация мониторинга, анализ данных.

Роль интеллектуальных транспортных систем как составной части транспортной инфраструктуры города, постоянно растет. Для поддержания качественного транспортного обслуживания транспортные предприятия активно внедряют автоматизированные системы управления, направленные на сокращение эксплуатационных расходов и упрощение диспетчерского управления транспортом. В последнее время, отдельным классом задач в таких системах становится помощь в принятии управленческих решений. Применительно к системам мониторинга и управления транспортом можно выделить множество задач, решение которых позволяет существенно повысить эффективность работы диспетчерской службы. В рамках данной статьи рассмотрим задачу прогнозирования развития дорожной обстановки на основе методов нечеткого поиска.

В целом региональную навигационно-информационную систему мониторинга и управления транспортом можно представить в виде набора взаимодействующих подсистем. Пример схемы функциональной структуры региональной навигационно-информационной системы представлен на рис. 1.

На основе агрегированной информации становится возможным решать оптимизационные задачи по управлению транспортными средствами, например, городским пассажирским транспортом. Для этого одним из ключевых моментов является своевременное прогнозирование развития дорожной обстановки. Рассмотрим данную задачу более подробно.

Поступающие данные о дорожной обстановке необходимо привести к единому виду (установление принадлежности данных к различным источникам поступления, анализ их типа). Из различных систем могут поступать одни и те же наборы параметров, однако их значения могут при этом отличаться. Например, информация о наличии пробок (затрудненном движении) поступает как от подчиненных транспортных средств, так и из внешних Web-ресурсов, однако из-за различных источников данных значения могут отличаться. Различаться могут и форматы данных, например, оценка дорожной обстановки может быть предоставлена в баллах, где минимум – нет затруднений для движения, а максимум – движение парализовано (или наоборот), или в виде средней скорости движения по улице.

После верификации полученных данных становится возможным сопоставить одинаковые параметры из разных источников, сравнить показатели за разные периоды, определяя тем самым корректность и достоверность данных, выбрать граничные данные для расчета оптимистичного/пессимистичного прогнозов.

На основе подготовленных данных с помощью поисковых алгоритмов становится возможным прогнозирование развития дорожной обстановки.

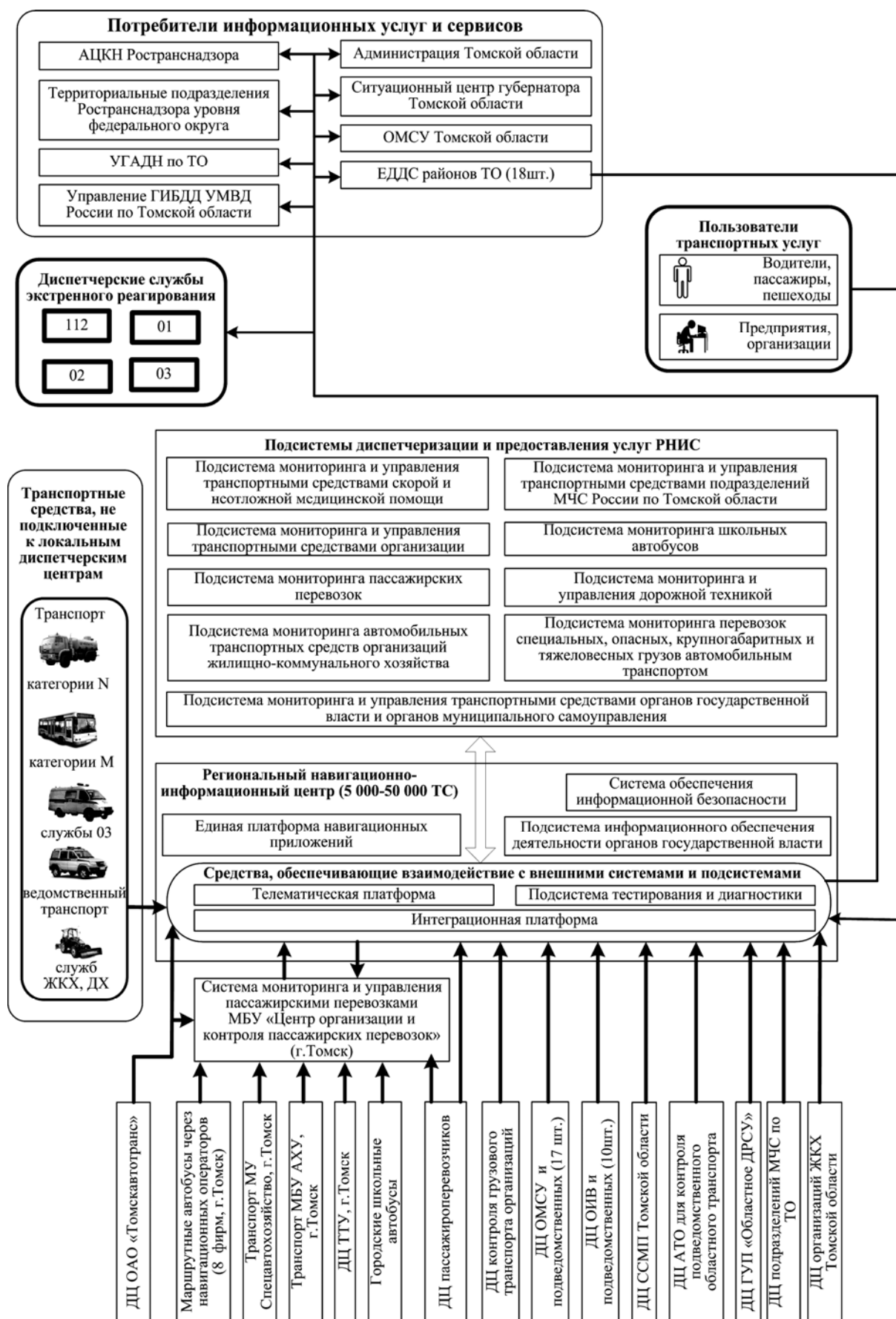


Рис. 1. Схема функциональной структуры региональной навигационно-информационной системы Томской области

Для прогнозирования любого процесса необходимо заранее определить ряд условий, выполнение которых с высокой вероятностью свидетельствует о наступлении контролируемого явления. Например, для прогнозирования развития дорожной обстановки необходимы исторические (статистические) данные по развитию дорожной обстановки с привязкой к времени наступления, динамике развития. Таким образом, задача прогнозирования развития дорожной обстановки может быть сведена к поиску в базе данных набора записей, соответствующих аналогичным условиям. В таком случае задача прогнозирования развития дорожной обстановки может быть решена с применением теории нечетких множеств (далее - на основе алгоритмов и методов нечеткого поиска).

К реализации нечеткого поиска в реляционных базах данных существует несколько подходов. Одним из таких подходов является нечеткий текстовый поиск, когда искомый объект и запись таблицы базы данных преобразуются в строки путем слияния атрибутов, а затем на основе алгоритмов нечеткого поиска строк вычисляется степень их сходства. Поиск по сходству подразумевает отыскание слов, для которых метрика (расстояние до поискового шаблона) не превышает заданную величину. Функция Левенштейна–Дамерау, наиболее популярная из существующих метрик, исчисляется минимальным количеством элементарных операций редактирования, необходимых для преобразования одной строки в другую, в том числе операции замены, вставки и удаления одного символа с учетом транспозиции символов.

На базе метрики Левенштейна–Дамерау построено большое число поисковых алгоритмов, одним из наиболее эффективных алгоритмов является метод N -грамм [1]. Данная разновидность нечеткого текстового поиска основана на определении функции нечеткого сравнения, которая 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 (1) будет выглядеть следующим образом:

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

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

где $\text{Count}(\text{Str}, i) = (\text{len}(\text{Str}) - i + 1)$; $\text{len}(S)$ – длина строки S ; $\text{Match}(S1, S2, i)$ = сумма совпадений всех подстрок длиной i из $S1$ в строке $S2$ [2, 3].

Нечеткие запросы также служат одним из подходов к реализации нечеткого поиска в реляционных базах данных. В основу нечетких запросов положена теория нечетких множеств Люфти Заде, где базовой характеристикой нечеткого множества выступает функция принадлежности $\mu(x)$, т.е. степень принадлежности элемента x к нечеткому множеству. Тогда нечетким множеством C называется множество упорядоченных пар вида $C = \mu(x)/x$, при этом $\mu(x)$ может принимать любые значения в интервале $[0, 1]$. Значение $\mu(x) = 0$ означает отсутствие принадлежности к множеству, 1 – полную принадлежность.

Например, треугольная функция (2) принадлежности определяется тройкой чисел (a, b, c) , и ее значение в точке x вычисляется согласно выражению [4]

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

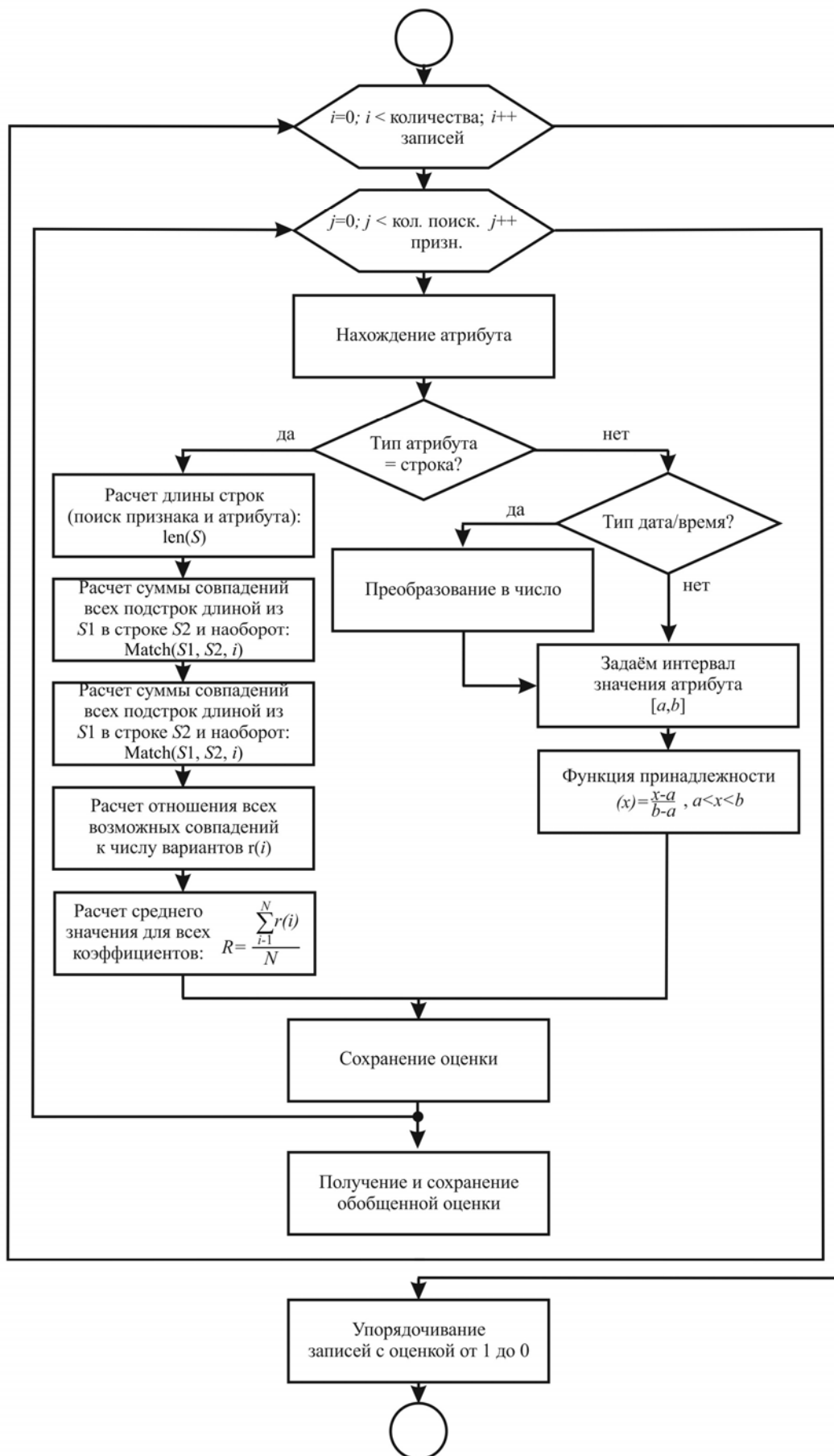


Рис. 2. Блок-схема комбинированного алгоритма прогнозирования развития дорожной обстановки

Рассмотрим указанные подходы применительно к задаче реализации нечеткого поиска для прогнозирования развития дорожной обстановки. В качестве алгоритма нечеткого текстового поиска будем использовать метод N -грамм. В данном методе 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. Применительно к строкам, полученным слиянием атрибутов записей таблиц базы данных, использование метода N -грамм будет неэффективным, поскольку помимо строковых значений записи таблиц базы данных содержат числовые значения, в отношении которых гораздо результативнее реализовать поиск с помощью нечетких запросов.

Для реализации нечетких запросов необходимо формализованно описать множества значений термов логического выражения. Далее также определяется функция принадлежности элемента множества (числового значения из записи таблицы базы данных) нечеткому множеству (заданному интервалу). 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 для прогнозирования развития дорожной обстановки предлагается использовать комбинированный алгоритм нечеткого поиска (рис. 2). Поскольку оба вышеизложенных подхода позволяют осуществлять поиск с использованием функций «сходства», 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 действие комбинированного алгоритма прогнозирования развития дорожной обстановки 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, 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. Затем 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. На основании полученных набора значений вычисляется обобщенная оценка принадлежности анализируемой записи объекту. 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) [5].

Прогнозирование развития дорожной обстановки позволяет существенно повысить эффективность диспетчерского управления пассажирским транспортом за счет оперативного управления транспортом.

Литература

1. Бойцов Л.М. Классификация и экспериментальное исследование современных алгоритмов нечеткого словарного поиска // Электронные библиотеки: перспективные методы и технологии, электронные коллекции: труды VI Всерос. науч. конф. (RCDL'2004). г. Пущино, 2004 [Электронный ресурс]. – Режим доступа: <http://www.rcdl.ru/papers/2004/paper27.pdf> (дата обращения: 20.05.2012).

2. Карахтанов Д.С. Использование алгоритмов нечеткого поиска при решении задачи устранения дубликатов в массивах данных // Молодой ученый. – 2010. – Т. 1, вып. 8(19). – С. 150–155.
3. Карахтанов Д.С. Использование алгоритмов нечеткого поиска при решении задач обработки массивов данных в интересах кредитных организаций // Аудит и финансовый анализ. – 2010. – Вып. 2 [Электронный ресурс]. – Режим доступа: <http://www.auditfin.com/2010/2/toc.asp> (дата обращения: 15.05.2012).
4. Потапов Е.Н. Нечеткие множества в хранилище данных // Проектирование корпоративных хранилищ данных. [Электронный ресурс]. – Режим доступа: <http://разработка-хд.рф/blog/?p=346> (дата обращения: 01.05.2012).
5. Рыжов А.П. Модели поиска информации в нечеткой среде. – М.: Изд-во Центра прикладных исследований при механико-математическом факультете МГУ, 2004. – С. 63–66.

Марков Николай Григорьевич

Д-р техн. наук, профессор, зав. каф. вычислительной техники
Национального исследовательского Томского политехнического университета (НИ ТПУ)
Тел.: (382-2) 70-17-77 доб. 22-22, (382-2) 61-20-71
Эл. почта: markovng@tpu.ru

Сонькин Дмитрий Михайлович

Канд. техн. наук, ассистент каф. информатики и проектирования систем НИ ТПУ
Тел.: (382-2) 42-67-27
Эл. почта: sonkind@tpu.ru

Газизов Тимур Тальгатович

Канд. техн. наук, зам. директора Института прикладной информатики
Томского государственного педагогического университета
Тел.: (382-2) 52-11-26
Эл. почта: gtt@tspu.edu.ru

Лещик Юлия Вадимовна

Аспирант каф. информатики и проектирования систем НИ ТПУ
Тел.: (382-2) 42-67-27
Эл. почта: catlen@ya.ru

Фадеев Александр Сергеевич

Канд. техн. наук, доцент каф. автоматизированных и компьютерных систем НИ ТПУ
Тел.: (382-2) 41-89-07
Эл. почта: fas@tpu.ru

Шемяков Александр Олегович

Канд. техн. наук, заместитель проректора Московского авиационного института по научной работе
Тел.: +7-910-423-84-92
Эл. почта: a.shemyakov@gmail.com

Markov N.G., Sonkin D.M., Gazizov T.T., Leschik Y.V., Fadeev A.S., Shemyakov A.O.

Combined road conditions prediction algorithm based on the methods of fuzzy search in the regional navigation-information system for monitoring and management of transport

With development of computer technology and increase in its performance, preparation of management decisions in automated systems becomes a separate class of problems. Applied to traffic management and monitoring systems can be identified a variety of tasks, automated solution that can significantly improve the efficiency of dispatching service. In this article consider the problem of road conditions prediction (traffic congestion prediction). On the basis of the known approaches fuzzy search in relational databases is proposed combined road conditions prediction algorithm.

Keywords: fuzzy set theory, fuzzy search algorithm, automation, monitoring, data analysis.