

ПЕТРЕНКО Т.Г., к.т.н., доцент,
 ТИМЧУК О.С., к.т.н., доцент,
 КУЦЕНКО Д.Д., магістр (ДонНУ, г. Вінниця)

Нечеткая муравьиная система для решения задачи о рюкзаке

В статье предложена нечеткая муравьиная система отбора информационных ресурсов. Выбор наиболее предпочтительной информации для пользователя построен на базе анализа нечетких параметров популярности информационных ресурсов.

Ключевые слова: нечеткая муравьиная система, модель рюкзака с мультिवыбором, текстовые ресурсы, популярность текста, гороскоп.

Введение

Задача о рюкзаке впервые была сформулирована для нахождения набора предметов, которые можно поместить в рюкзак ограниченной вместимости [1]. Варианты данной задачи содержат дополнительные ограничения на типы рюкзаков, наборы предметов и критерии их отбора. Для решения задачи о рюкзаке использованы различные алгоритмы отбора набора предметов (ресурсов) [2]. Задача о рюкзаке относится к классу NP-полных задач, поэтому масштабирование выбираемых ресурсов приводит к повышению затрат на решение. Использование эволюционных алгоритмов в таком случае позволяет существенно снизить размерность задач, но приводит к получению приближенного результата [3]. Использование нечеткой логики наряду с эволюционными алгоритмами приводит к дальнейшему снижению размерности [4], а также позволяет включить в обработку информацию, которая представляется в системе нечетко. Примером решения таких задач является обработка web-ресурсов в on-line режиме [5-7]. Актуальным является подход, который объединяет преимущества традиционных подходов к решению задачи о рюкзаке с помощью алгоритма муравьиной колонии и преимущества нечетких систем 2-го типа для обработки данных, характеристики которых представлены нечетко.

Постановка задачи

Для информационных ресурсов, которые размещены на разных сайтах и группируются в некоторые категории, разработать систему отбора информационных ресурсов для их предоставления заинтересованным пользователям. Систему отбора ресурсов организовать как нечеткую систему 2-го типа с ограничениями и критериями задачи о рюкзаке. Для оптимизации отбора ресурсов использовать муравьиный алгоритм.

Решение задачи

В данной работе в качестве информационных ресурсов рассматривается текстовая информация.

Сформулируем задачу о рюкзаке с мультिवыбором:

$$\sum_i \sum_j p_{ij} x_{ij} \rightarrow \max, i = \overline{1, I}, j \in S_i,$$

$$\sum_i \sum_j w_{ij} x_{ij} \leq W,$$

$$x_{ij} = \begin{cases} 0, & \text{информация не помещается в рюкзак,} \\ 1, & \text{информация помещается в рюкзак,} \end{cases}$$

где I – количество источников – web-ресурсов, которые предоставляют текстовую информацию,

S_i – выбранные классы (категории) текстовой информации из i -го источника,

p_{ij} – популярность текстовой информации,

x_{ij} – признак помещения текстовой информации в рюкзак,

w_{ij} – объем текстовой информации,

W – объем рюкзака.

Необходимо получить такой набор информации, который бы включал все категории информации и отвечал требованию тематического соответствия текстов некоторому эталону, а значит, решал бы проблему выбора наиболее подходящего источника информации для каждой категории информации.

В работе для решения задачи о рюкзаке используется АСО алгоритм [8]. Алгоритм построения графа для перемещения муравьев для задачи о рюкзаке представлен в работе [9]. Удельная ценность категории информации η_{ij} , которая выражает эвристическое желание муравья взять текстовую информацию, в данной работе определяется как количество баллов, выставленных пользователями при голосовании. Для определения лучшего результата выполняется оценка

ценности – популярности виртуальных рюкзаков муравьев.

Оценка популярности рюкзака k -го муравья строится на базе анализа и обработки параметров популярности информации. При анализе предметной области было установлено, что параметры популярности являются неопределенными. Основные причины неопределенности:

- интуитивный характер параметров;
- словесное описание параметров;
- отсутствие однозначной интерпретации параметров среди разработчиков программного обеспечения;
- отсутствие согласованных методов и инструментов обработки параметров;
- непостоянство параметров в течение определенного промежутка времени.

В данной работе для учета неопределенности такого рода предлагается использовать методы теории дискретных интервальных нечетких систем второго типа (DIT2FLSs) [10]. В соответствии с положениями теории DIT2FLSs представим модель оценки популярности текстовой информации:

$$y = F(IN_k, LI, LO, R),$$

$$IN_k = \langle in_n \rangle, n = \overline{1, N},$$

$$LI = \langle li_n \rangle,$$

$$R = \langle r_m \rangle, m = \overline{1, M},$$

где IN_k – набор четких входных значений (параметры популярности текстовой информации),

k – номер муравья,

N – количество входных значений,

LI – набор входных лингвистических переменных второго типа, которые описывают параметры популярности текстовой информации,

LO – результирующая лингвистическая переменная второго типа, описывающая уровень популярности текстовой информации,

R – набор нечетких правил, на основе которых выполняется оценка популярности текстовой информации,

M – количество нечетких правил,

F – операция нечеткого логического вывода (алгоритм Мамдани),

y – уровень популярности текстовой информации.

Популярность категории информации для пользователя зависит от многих параметров, однако самыми существенными являются – важность и объем

где in_1 – объем текстовой информации; значение in_1 приводится к нормированному интервалу $[0; 1]$,

in_2 – важность текстовой информации, которая определяется как отношение количества ключевых слов категории k общему количеству слов в тексте определенной категории; значение in_2 приводится к нормированному интервалу $[0; 1]$.

Набор LI содержит 2 лингвистические переменные

$$LI = \langle li_1, li_2 \rangle,$$

где li_1 – лингвистическая переменная, описывающая объем информации; содержит 3 терма; определена на универсальном множестве $X_1 = [0; 1]$ (рис. 1 а),

li_2 – лингвистическая переменная, описывающая уровень важности информации; содержит 3 терма; определена на универсальном множестве $X_2 = [0; 1]$ (рис. 1 б).

Результирующая лингвистическая переменная LO содержит 3 терма; определена на универсальном множестве $X_{res} = [0; 1]$ (рис. 1 в).

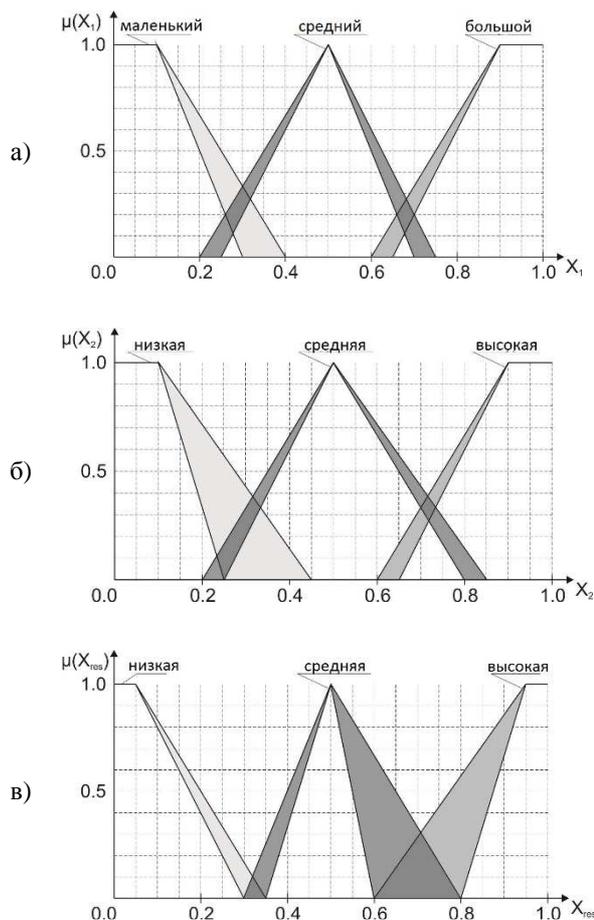


Рис. 1. Лингвистические переменные модели оценки популярности текстовой информации: (а), (б) – входные, (в) – результирующая.

$$IN(t) = \langle in_1, in_2 \rangle,$$

Набор правил R содержит 9 нечетких правил типа IF-THEN (табл. 1). Для каждого правила экспертом установлена степень доверия к правилу.

Таблица 1

База правил модели оценки популярности текстовой информации

| l_i \ l_j | маленький | средний | большой |
|---------------|---------------|---------------|---------------|
| низкая | средняя (0.6) | низкая (0.9) | низкая (1.0) |
| средняя | средняя (0.9) | средняя (0.8) | низкая (0.5) |
| высокая | высокая (1.0) | высокая (1.0) | средняя (1.0) |

Программная реализация

Система отбора ресурсов была использована в программе «Гороскоп на завтра» для пользователей мобильных телефонов [11]. Отбор ресурсов выполняется в серверной части программы, а клиентское приложение устанавливается на мобильный телефон. Система отбора ресурсов написана с применением языков программирования PHP и Java, технологии Ajax и библиотеки JQuery. Клиентское приложение реализовано на платформе Android.

Система отбора текстовой информации в приложении «Гороскоп на завтра», построенная как нечеткая муравьиная система, решает задачу о рюкзаке и позволяет расширять ресурсы системы. Возможно добавление новых категорий информации (гороскопов) и новых нечетких параметров в систему отбора информации из различных источников свободного доступа. Приложение «Гороскоп на завтра» обладает высоким рейтингом среди аналогичных программ: количество ежедневных обращений к программе «Гороскоп на завтра» составляет около 500; программа установлена 22000 раз (примерно 100 установок за день); оценка программы пользователями составляет в среднем 4.3 балла из 5 (проголосовало 463 пользователя); выполняется ежедневное обновление сервисов программы (отбор информации); в текущей версии программа предоставляет 4 категории информации – гороскопы любовный, карьерный, гороскоп здоровья и общий гороскоп.

Выводы

Задача о рюкзаке имеет 25-ти летнюю историю и множество реализованных на практике приложений. Однако, не менее успешна 50-ти летняя история развития нечеткой логики. Объединение преимуществ различных направлений позволяет формировать гибридные модели с новыми возможностями. В данной работе предложен подход, который нечеткую муравьиную систему ограничивает рамками задачи о рюкзаке. Такой подход позволил создать систему

отбора информационных ресурсов, которые обладают нечеткими характеристиками. Рюкзак ограниченной вместимости заполняется обязательными категориями информации (набором категорий). Формирование вариантов набора выполняется муравьиной колонией (параметры муравьиного алгоритма подобраны экспериментально), а оценка каждого набора выполняется с помощью нечеткой системы 2-го типа. Применение предложенного в работе подхода к отбору текстов гороскопов из различных источников и формирование набора гороскопов показало хорошие результаты в случае масштабирования ресурсов.

Литература

1. S. Martello, P. Toth Knapsack problems: algorithms and computer implementations. Wiley&Sons, Chichester, 1990.
2. H. Kellerer, U. Pferschy, D. Pisinger Knapsack Problems. Springer Science & Business Media, 2004.
3. Bo Xing, Wen-Jing Gao Innovative Computational Intelligence: A Rough Guide to 134 Clever Algorithms. Intelligent Systems Reference Library. Volume 62. Springer, 2014, [Электронный ресурс] Режим доступа: <http://www.springer.com/series/8578> - Дата доступа: 6.09.2015.
4. J. Casillas, O. Cordon, F. Herrera, Learning fuzzy rules using ant colony optimization algorithms, Proceedings of the 2nd International Workshop on Ant Algorithms (ANTS 2000), Brussels, Belgium (2000) 13-21. [Электронный ресурс]. Режим доступа: <http://decsai.ugr.es/~casillas/public.php> - Дата доступа: 6.09.2015.
5. D. Chakrabarty, Y. Zhou, R. Lukose Online Knapsack Problems. [Электронный ресурс]. Режим доступа: <http://research.microsoft.com/en-us/um/people/dechakr/pubs/czk-full.pdf> - Дата доступа: 6.09.2015.
6. D. Pisinger A Minimal Algorithm for the Multiple-Choice Knapsack Problem, 1994, [Электронный ресурс]. Режим доступа: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.11.4976&rep=rep1&type=pdf> - Дата доступа: 6.09.2015.
7. M. Cygan, Ł. Jez, J. Sgall Online Knapsack Revisited, 2014, [Электронный ресурс]. Режим доступа: <http://link.springer.com/article/10.1007%2Fs00224-014-9566-4> - Дата доступа: 6.09.2015.
8. Dorigo M. Ant Colony Optimization / M. Dorigo, T. Stützle – MIT Press, 2004. – 319 p.
9. Кагиров Р.Р. Многомерная задача о рюкзаке: новые методы решения / Р.Р. Кагиров // Вестник Сибирского государственного аэрокосмического университета им. академика М.Ф. Решетнева. – 2007. – № 3. – С. 16-20.

10. Petrenko T. Package Library and Toolbox for Discrete Interval Type-2 Fuzzy Logic Systems / Т. Petrenko, О. Tymchuk // Proceedings of the 18th International Conference on Soft Computing (MENDEL), 2012, June 27-29, Brno, Czech Republic. – pp. 233 – 238.
11. Приложение «Гороскоп на завтра», [Электронный ресурс]. Режим доступа: <https://play.google.com/store/apps/details?id=ua.maksdenis.horoscope> - Дата доступа: 6.09.2015.

Петренко Т.Г., Тимчук О.С., Куценко Д.Д. Нечітка мурашина системи вирішення задачі про рюкзак. У статті запропонована нечітка мурашина система відбору інформаційних ресурсів. Вибір найбільш кращою інформації для користувача побудований на базі аналізу нечітких параметрів популярності інформаційних ресурсів.

Ключові слова: нечітка мурашина система, модель рюкзака з мультिवибором, текстові ресурси, популярність тексту, гороскоп.

Петренко Т.Г., к.т.н., доцент кафедри комп'ютерних технологій, Донецький національний університет, Вінниця, Україна

Тимчук О.С., к.т.н., доцент кафедри комп'ютерних технологій, Донецький національний університет, Вінниця, Україна

Куценко Д.Д., магістр з кваліфікацією «Розробник програмного забезпечення», Донецький національний університет, Вінниця, Україна

Petrenko T., Tymchuk O., Kutsenko D. Fuzzy ant system to solve the knapsack problem. This paper presents the information resources selection system, in which the resources are represented as text information with fuzzy parameters. The resources are grouped according to categories. The resources selection system is based on the algorithm of ant colony and type-2 fuzzy system. The selection of text information is limited by the restrictions of knapsack problem – a knapsack of bounded capacity is filled with the required set of categories of text information. Ant colony generates variants of category sets. Type-2 fuzzy system evaluates the ant's virtual knapsack. The evaluation is built on the analysis of fuzzy parameters of text information popularity (the volume and importance of information).

The proposed resources selection system has been implemented in the application "Horoscope for Tomorrow" (Google play Resource). It offers a set of horoscopes as text information categories to mobile phone users. The selection of categories is made from various web-sources. The application of the proposed resources selection system to the formation of horoscope sets has given good results: the number of active users is approximately 500; 22000 installations.

Key words: fuzzy ant system, multiple-choice Knapsack model, text resources, text popularity, horoscope.

Рецензент д.т.н., професор Каргин А.А. (ДонНУ, г. Вінниця)

Поступила 08.09.2015г