УДК 519.237.8

КЛАСТЕРИЗАЦИЯ ПОЛЬЗОВАТЕЛЕЙ ИНФОРМАЦИОННОЙ СИСТЕМЫ НА ОСНОВЕ ИХ ДЕЙСТВИЙ В КОМПЬЮТЕРНОЙ СЕТИ

[1]Назаров А.О.

[1]ФГБОУ ВПО «Казанский национальный исследовательский технический университет им. А.Н. Туполева-КАИ», Казань, Россия (420111, г. Казань, ул. К. Маркса, 10), e-mail:sas4406@yandex.ru

В статье предложен процесс кластеризации пользователей на основе их действий в информационной системе. Разработан метод концептуальной кластеризации пользователей информационной системы, который формирует пользовательские роли на основе действий пользователей в информационной системе. Разработанный метод способен работать с объектами, характеризуемыми множеством нечетких параметров и основан на классическом методе концептуальной кластеризации COBWEB. Описан программный комплекс, реализующий разработанный метода концептуальной кластеризации. Решена практическая задача кластеризации пользователей на основе реальной информационной системы конкретной организации. Решение данной задачи позволяет с одной стороны значительно упростить работу администратора информационной безопасности по формированию пользовательских ролей в информационной системе, с другой стороны позволяет обнаруживать аномальное поведение пользователей в компьютерной сети.

Ключевые слова: информационная система, пользовательские роли, кластеризация данных.

CLUSTERING USER INFORMATION SYSTEM BASED ON THEIR ACTION COMPUTER NETWORK

[1]Nazarov A.O.

[1]A. Tupolev Kazan State Technical University - KAI , Kazan, Russia (420111, Kazan, street K.Marksa, 10), e-mail:sas4406@yandex.ru

This article proposes a process of clustering users based on their actions in the information system of the subject. Developed a method for conceptual clustering, which generates custom roles based on user actions in the information system. The developed method is able to work with objects, characterized by a plurality of fuzzy parameters and is based on the classical method of conceptual clustering COBWEB. Describes a software package that implements the method developed ethyl conceptual clustering. Solved the practical problem of clustering users based on real information system specific organization. The obtained results allows users to allocate poured characterized by abnormal behavior of a computer network.

Keywords: information system, user roles, data clustering.

В информационных системах часто возникают инциденты, приводящие к проблемам информационной безопасности. Например, это могут быть пользователи, занимающиеся не своими обязанностями. Но еще опаснее, если эти пользователи нелегитимны и под пользователем аутентифицировался злоумышленник. В связи с этим, существуют необходимость в выявлении и пресечении данных фактов. В качестве одного из способа решения данной проблемы, является внедрение политики разграничения доступа к информационным ресурсам организации на основе ролевого разграничения доступа [1].

Грамотное разграничение доступа пользователей к ресурсам информационной системы максимально снижает возможность выполнения нелегитимных действий со стороны пользователей. В современных системах разграничение доступа достигается за счет использование пользовательских ролей [2].

Формирование ролей в системах с большим количеством пользователей и прав доступа является сложной, ресурсоемкой задачей, требующей больших временных затрат. Существуют различные методы автоматизация данного процесса [4, 6]. Наиболее оптимальным решением является использование методов кластеризации.

Интуитивно понятно, что применение метода кластеризации данных в качестве вспомогательного средства создания ролей заключается в том, что каждому пользователю, действующему в рамках роли, присущ доступ к некоторым приложениям и комбинация этих прав неявно указывают на роль. Предполагается, что значительное число пользователей, имеющих одинаковый набор прав, объединяются в роль.

Набор прав является достаточно четким параметром, так как его можно описать четкими значениями, т.е. либо данное право у пользователя есть, либо его нет. Сложнее описать четкими значениями действия пользователей в ИС. В связи с этим, для кластеризации объектов, описанных нечеткими параметрами, в последнее время активно используются методы нечеткой кластеризации.

Целью исследования является разработка метода кластеризации пользователей, на основе их поведения в информационной системе. Предлагается разработать новый метод кластеризации на основе существующего метода кластеризации COBWEB. Эффективность разрабатываемого метода концептуальной кластеризации определяется его способностью проводить кластеризацию пользователей информационной системы, на основе их действий в информационной системе.

Материалы и методы исследования

В статье предложен метод нечеткой концептуальной кластеризации, основанный на методе COBWEB, позволяющий работать с нечеткими объектами. Данный метод предполагает реализацию классического метода концептуальной кластеризации [3] в следующих условиях:

1. Множество распознаваемых объектов $O = \{O_i\}_{i=\overline{1,r}}$, характеризуется нечеткими параметрами $\tilde{A} = \{\tilde{A}_j\}_{j=\overline{1,m}}$,

2. Значение параметра $\tilde{A}_j$ для объекта O_i определяется в виде функции принадлежности $\mu_{\tilde{A}_{ij}}(x) \in \{0;1\}$.

3. Степень сходства двух функций принадлежности $\mu_{\tilde{A}_{ij}}(x)$ и $\mu_{\tilde{A}_{ij}}(x)$ определяется их наибольшей верхней границей, в виде:

$$v_{jit} = \sup_{x \in X} \min_{x \in X} \{\mu_{\tilde{A}_{ij}}(x), \mu_{\tilde{A}_{ij}}(x)\} \in [0,1], \quad (1)$$

где $\mu_{\tilde{A}_{ij}}(x)$ - функция принадлежности параметра $\tilde{A}_j$ для объекта O_i, а $\mu_{\tilde{A}_{tj}}(x)$ - функция принадлежности параметра $\tilde{A}_j$ для объекта O_t.

4. Основываясь на формуле полезности кластеризации классического метода концептуальной COBWEB [3] и условиях 1-3, оценка полезности кластеризации осуществляется по модифицированной формуле (2)

$$CU^* = \frac{\sum_{k=1}^{n} P(C_k)\left[\sum_{j=1}^{m}\sum_{i=1,O_i\in C_k}^{r}\sum_{t=1,O_t\in C_k}^{r} v_{jit}/|C_k| - \sum_{j=1}^{m}\sum_{i=1}^{r}\sum_{t=1}^{r} v_{jit}/r\right]}{n}, \qquad (2)$$

где n – количество кластеров.

Для практического решения задач на основе предложенного метода нечеткой концептуальной кластеризации, был разработан программный комплекс в среде C#.

Архитектура системы, предназначенной для автоматизации построения пользовательских ролей, представлена на рисунке 1 и состоит из 4 модулей.

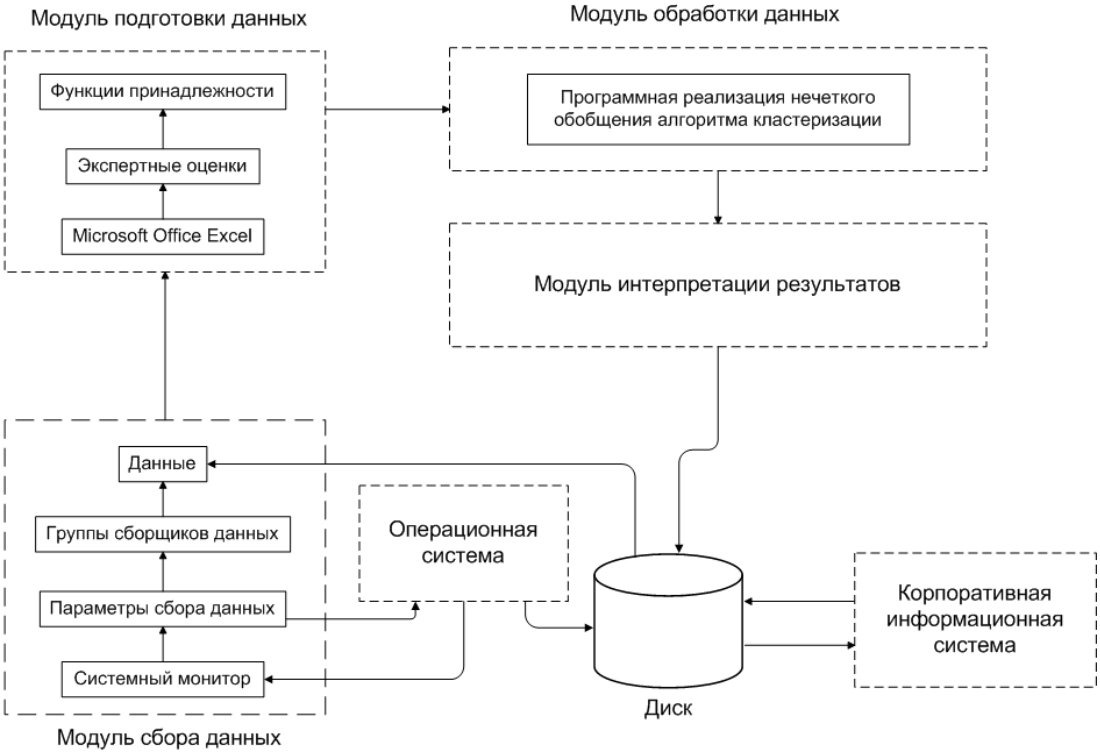

Рис. 1. Архитектура системы, предназначенной для автоматизации построения пользовательских ролей

1. Модуль сбора данных, предназначенный для сбора статистических данных об объектах. Для решения задачи кластеризации пользователей в ИС сбор данных осуществляется на основе анализа журналов событий. Статистические данные сохраняются на жестком диске в виде текстового файла с разделителями.

2. Модуль подготовки данных на основе полученных статистических данных, с привлечением эксперта, позволяет построить функции принадлежности для каждого объекта по каждо-

му из параметров. Выходом данного модуля являются сформированные нечеткие описания объектов в виде функций принадлежности их параметров.

3. Модуль обработки данных реализует разработанный метод нечеткой концептуальной кластеризации.

Результаты работы метода можно увидеть в отдельном окне, в котором представляются результаты кластеризации и полученная иерархия кластеров.

Результаты исследования и их обсуждения

В качестве примера задача автоматизации формирования пользовательских ролей была решена на действующей информационной системе конкретной организации. Структура информационной системы представлена на рисунке 2. Она включает в себя четыре отдела: Администрация, Бухгалтерия, Отдел продаж, Технический отдел. В состав ЛВС входит один почтовый и два файловых сервера, коммутатор, сетевой принтер, 3 МФУ. Осуществляется выход во внешнюю сеть через прокси-сервер.

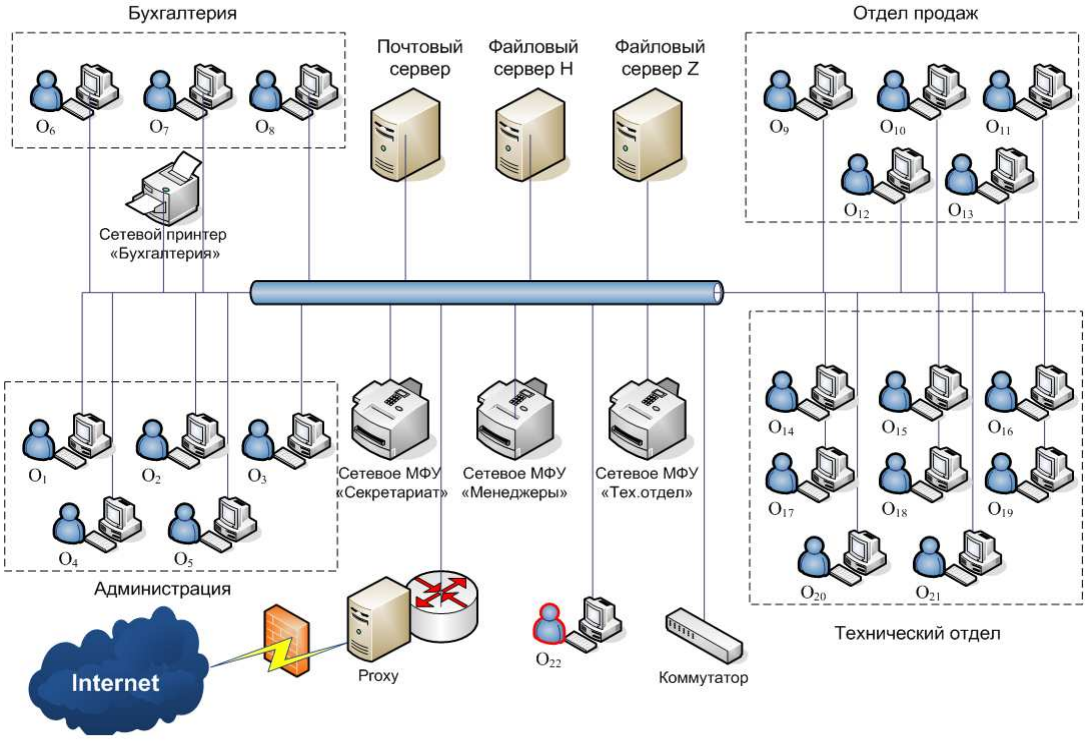

Рис. 2. Структура информационной системы организации

Осуществлялась кластеризация 22 пользователей: $O = \{O_i\}_{i=1}^{22}$, представленных в таблице 1.

Таблица 1

Пользователи информационной системы

Пользователь	Должность	Отдел
O1	Генеральный директор	Администрация

O_2	Финансовый директор	
O_3	Технический директор	
O_4	Секретарь	
O_5	Офис-менеджер	
O_6	Главный бухгалтер	Бухгалтерия
O_7	Зам. главного бухгалтера	
O_8	Бухгалтер	
O_9	Начальник отдела продаж	Отдел продаж
O_{10}	Зам. начальника отдела продаж	
O_{11}- O_{13}	Менеджеры	
O_{14}	Начальник технического отдела	Технический отдел
O_{15}	Зам. начальника технического отдела	
O_{16} – O_{21}	Технические специалисты	
O_{22}	Администратор ЛВС	Отдел ИТ

Для описания поведения пользователей было выделено 18 параметров $A = \{A_j\}_{j=1}^{18}$, представленных в таблице 2.

Таблица 2

Параметры пользователей

Параметр	Описание параметра
$\tilde{A}_1$	Количество обращений к почтовому серверу в сутки
$\tilde{A}_2$	Количество обращений к файловому серверу H в сутки
$\tilde{A}_3$	Количество обращений к файловому серверу Z в сутки
$\tilde{A}_4$	Количество обращений к коммутатору в сутки
$\tilde{A}_5$	Количество обращений к сетевому принтеру «Бухгалтерия» в сутки
$\tilde{A}_6$	Количество обращений к сетевому МФУ «Секретариат» в сутки
$\tilde{A}_7$	Количество обращений к сетевому МФУ «Менеджеры» в сутки
$\tilde{A}_8$	Количество обращений к сетевому МФУ «Тех.отдел» в сутки
$\tilde{A}_9$	Количество обращений к прокси-серверу в сутки
$\tilde{A}_{10}$	Количество принятых, отправленных писем через Microsoft Office Outlook в сутки
$\tilde{A}_{11}$	Количество обращений к «1С:Бухгалтерия 8» в сутки
$\tilde{A}_{12}$	Количество обращений к «1С:Документооборот» в сутки

$\tilde{A}_{13}$	Количество обращений к «1С:Предприятие 8» в сутки
$\tilde{A}_{14}$	Количество обращений к «Microsoft Navision 3.60» в сутки
$\tilde{A}_{15}$	Количество обращений к «1С:Зарплата и управление персоналом 8» в сутки
$\tilde{A}_{16}$	Количество обращений к модулю «Монитор сопровождения» в сутки
$\tilde{A}_{17}$	Объем внешнего сетевого трафика в сутки
$\tilde{A}_{18}$	Средний процент загруженности центрального процессора в сутки

На основании анализа поведения пользователей по выше перечисленным параметрам, осуществлялась кластеризация и распределение пользователей по кластерам. Результаты кластеризации представлены в таблице 3.

Таблица 3

Результаты проведенной кластеризации

Кластер	Объект (Пользователь)
C_1	O_1
C_2	O_2
C_3	O_3
C_4	O_4, O_5
C_5	O_6, O_7, O_8
C_6	O_9, O_{10}, O_{11}, O_{13}
C_7	O_{12}
C_8	O_{14}, O_{15}, O_{16}, O_{18}, O_{19}, O_{20}, O_{21}
C_9	O_{17}
C_{10}	O_{22}

Анализируя таблицу 3, можно сделать вывод, что программа сформировала 10 кластеров. Первые три кластера описывают действия пользователей руководящего состава генерального директора, финансового директора и технического директора соответственно. Кластер C_4 описывает поведение объектов O_4 и O_5 – пользователи секретарь и офис-менеджер.

В связи с функциональными обязанностями администратора ЛВС, его действия в ИС отличны от действий других пользователей. Поэтому, для администратора ЛВС (O_{22}) был создан отдельный кластер.

Так же были выделены кластеры, характеризующие поведение пользователей, входящих в различные структурные подразделения организации:

Бухгалтерия – C_5;

Отдел продаж – C_6;

Технический отдел – C_8.

Пользователи, относящиеся к кластерам C_7 и C_9 были выделены в отдельные кластеры, что свидетельствует об их аномальном поведении. Проведя детальный анализ, было выявлено, что объем внешнего сетевого трафика объекта O_{12} превышает показатели использованного трафика других пользователей отдела продаж, что и формирует подобную аномалию. А пользователь O_{17} обращался к программам, использование которых не является необходимым при выполнении функциональных обязанностей сотрудника технического отдела, что также сформировало соответствующую аномалию. Выявление подобных инцидентов позволяет администратору безопасности своевременно отреагировать на них.

Заключение

С помощью разработанного метода нечеткой концептуальной кластеризации решена практическая задача по автоматизации построения пользовательских ролей в информационной системе. Решение данной задачи позволяет с одной стороны значительно упростить работу администратора информационной безопасности по формированию пользовательских ролей в информационной системе, с другой стороны позволяет обнаруживать аномальное поведение пользователей в компьютерной сети, выявляя недобросовестных сотрудников, использующих информационные ресурсы организации не только для выполнения своих функциональных обязанностей, но и в личных целях.

Список литературы

1. Гайдамакин Н.А. Разграничение доступа к информации в компьютерных системах. – Екатеринбург: изд-во Урал. Ун-та, 2003 г. – 328 с.
2. Девянин П.Н. Модели безопасности компьютерных систем: Учеб. пособие для студ. высш. учеб. заведений. — М.: Издательский центр «Академия», 2005. — 144 с.
3. Fisher D. Knowledge Acquisition Via Incremental Conceptual Clustering, 1987. – P. 142-153.
4. Han J., Pei J., Yin Y., and Mao R. Mining frequent patterns without candidate generation: A frequent-pattern tree approach. Data Mining and Knowledge Discovery, 2004. – P. 235-247.
5. Sato M., Sato Y., and Jain L. Fuzzy Clustering Models and Applications, Physica-Verlag, Heidelberg, 1997. – P. 135-148.
6. Zhang D. Data Mining for Role Based Access Control. Master-to-PhD Conversion Report, 2006. – P. 348.

Рецензенты:

Райхлин В.А., д.ф-м.н., профессор, кафедры «Компьютерных систем», ФГБОУ ВПО «КНИТУ им. А.Н.Туполева-КАИ», г. Казань.
Захаров В.М., д.т.н., профессор кафедры «Компьютерных систем», ФГБОУ ВПО «КНИТУ им. А.Н.Туполева-КАИ», г. Казань.