

УДК 004.89:519.7

DOI:

Мельников О. Ю., Спориш Д. К.**РОЗРОБКА СИСТЕМИ ДЛЯ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ ІНТЕРНЕТ-ПРОВАЙДЕРА**

Інтернет-провайдери – це компанії, що надають послуги доступу до всесвітньої мережі Інтернет. Ріст числа користувачів мережі Інтернет приводить до росту пропозиції в цій ніші й посиленню конкуренції між провайдерами. Кожен користувач, крім цілодобового високошвидкісного доступу до інтернет-сайтів, хоче мати й сервісну підтримку, у тому числі – у вигляді «особистого кабінету» на сайті провайдера з можливістю контролю балансу й платежів.

Багато виробників програмного забезпечення підтримують розробки так званих білінгових систем, призначених для обліку й тарифікації всього спектра послуг, надаваних операторами зв'язку (Dialup, VPN, Hotspot, VoIP, IPTV). Як правило, такі системи є багатофункціональними, модульними АСР (Автоматична Система Розрахунків) з відкритим програмним кодом, що забезпечують комплексні можливості в реальному масштабі часу, значну гнучкість і масштабованість для менеджменту й прискореного впровадження послуг. Це дозволяє операторові зв'язку розробляти нові послуги, ефективно управляти тарифами, користувачами й здійснювати різні операції, необхідні для оперативного реагування у відповідь на вимоги, що змінюються, ринку й запити споживачів [1-4].

Однак у більшості таких систем не підтримується використання математичних методів для інтелектуального аналізу наявних даних, що могло б краще обґрунтувати прийняті рішення. Термін Data Mining – інтелектуальний аналіз даних – часто розшифровується як видобуток даних, добування інформації, засобу пошуку закономірностей, добування знань, аналіз шаблонів, процес підтримки прийняття рішень, заснований на пошуку в даних схованих закономірностей (шаблонів інформації) [5-7].

У зв'язку з тим, що існує велика кількість методів для рішення низки задач інтелектуального аналізу даних, стає актуальним вибір таких методів, які повинні бути максимально ефективними для аналізу даних конкретного підприємства. Метою роботи є удосконалення процесу прийняття управлінських рішень на підприємстві «Інтернет-провайдер Matrix.Net» [8] шляхом дослідження методів інтелектуального аналізу даних та розробки інформаційної системи для проведення цього аналізу.

Білінгова система, використовувана провайдером «Matrix.Net», дозволяє вести облік і тарифікацію всього спектра надаваних послуг, забезпечує комплексні можливості в реальному масштабі часу. Однак у даній системі відсутня будь-яка інтелектуальна обробка наявних даних. Тому перспективним і доцільним можна вважати створення модуля (а точніше – окремої програмної системи) для інтелектуального аналізу наявних даних, а саме:

- даних про боржників (ПІБ, адреса, сума боргу, дата останньої сплати тощо);
- даних про дії операторів (дата та час, логін, дія тощо).

Вочевидь, що значимими є дані про район, суму боргу ($\text{Борг} = \text{Кредит} - |\text{сума}|$) і числі днів, що пройшли від дня останнього підключення. Поля логін, ПІБ, телефон, адреса та прімітка можна вважати інформаційними і до розгляду не приймати.

Для дій операторів значимими є дані про час дії й змісту цієї дії, інші поля можна вважати інформаційними та до розгляду не приймати. Однак значним фактором у другому випадку можна вважати не тільки вид дії оператора в якийсь момент часу, а ще й частоту виконання кожної дії за період часу.

Як відомо, виділяють десять задач інтелектуального аналізу даних [5–7], кожна з яких може бути вирішена низкою методів.

Перша задача, яку можна сформулювати у нашій предметній галузі – класифікація (або прогнозування) дій користувача. Нам цікаво, яку дію може здійснити то чи інший

користувач відповідно до часу. Цю задачу можна вирішити методом побудови дерева рішень [9]. Метод дерев рішень часто називають «наївним» підходом. Але завдяки цілому ряду переваг, даний метод є одним з найбільш популярних для вирішення задач класифікації. Проведені розрахунки для нашого прикладу у середовищі «Deductor Studio Lite» [10] свідчать, що атрибут «Час» є максимально впливовим, а атрибут «Оператор» аж ніяк не впливає на вибір. Тому отримані результати навряд чи можливо вважати вагомими.

Розглянемо інший метод рішення задачі класифікації – шляхом побудови штучної нейронної мережі [11-13]. Параметрами вхідного шару нашої нейронної мережі будуть час здійснення якоїсь дії (неперервне значення) та прізвище (логін) користувача, який здійснив цю дію. Вихід – саме опис дії. Отримані результати можуть свідчать про більшу точність розрахунків, аніж при використанні метода дерев рішень, але їх практична цінність викликає сумнів.

Більш значну практичну цінність може мати рішення задачі кластеризації. Постановку завдання можна описати в такий спосіб:

1. Класифікувати користувачів, що мають заборгованість (таблиця «Боржники») – об'єднати їх у групи для зручності відпрацьовування й висування гіпотез про причини виникнення боргу.

2. Класифікувати операторів залежно від переважних у їхній діяльності типах дій (таблиця «Дії оператора») – об'єднати їх у групи для зручності аналізу діяльності й виявлення переваг.

В обох випадках ми не можемо заздалегідь виділити число й характерні ознаки кожної групи (класу об'єктів), тому необхідно вирішувати завдання саме кластеризації, а не класифікації. Як відомо, кластеризація призначена для розбивки сукупності об'єктів на однорідні групи (кластери або класи). Кластеризація є описовою процедурою, вона не робить ніяких статистичних висновків, але дає можливість провести розвідницький аналіз і вивчити «структуру даних» [5-7].

Суть ієрархічної кластеризації складається в послідовному об'єднанні менших кластерів у більші або поділі більших кластерів на менші. При великій кількості спостережень використовують неієрархічні методи, засновані на поділі, які являють собою ітеративні методи дроблення вихідної сукупності. Самим поширеним є метод k-середніх [5, 14].

Таким чином, треба розробити діючу інформаційну систему – застосування, яке дозволяло б проводити розподіл об'єктів по класах, групувати боржників та операторів для подальшого прийняття рішень.

Оскільки ми не можемо заздалегідь чітко визначити число кластерів, у нашому випадку доцільне початкове використання алгоритму ієрархічної агломеративної кластеризації. Попередня обробка даних для рішення другого завдання повинна містити підрахунок числа факторів (видів дії оператора) та підрахунок кількості кожної дії кожного оператора. Далі можливо використання ітеративного метода k-середніх.

Інформаційна модель системи створюється уніфікованою мовою моделювання UML – Unified Modeling Language [15-16].

Діаграма варіантів використання проекрованої системи представлена на рис. 1. Для спрощення не будемо ділити користувачів по рівнях: єдиний актор є одночасно й адміністратором, і користувачем, і системним аналітиком. Користувачеві доступні дві основних дії – «Робота з боржниками» й «Робота з діями операторів». «Робота з боржниками» припускає обов'язкове завантаження вихідних даних по боржниках, обробку цих даних і проведення кластеризації, при чому користувач повинен обрати метод для проведення кластеризації. Також користувач може провести аналіз результатів кластеризації боржників.

«Робота з діями операторів» припускає обов'язкове завантаження вихідних даних по діях операторів, обробку цих даних і проведення кластеризації, при чому користувач також повинен обрати метод для проведення кластеризації. При необхідності користувач може провести аналіз результатів кластеризації дій операторів.

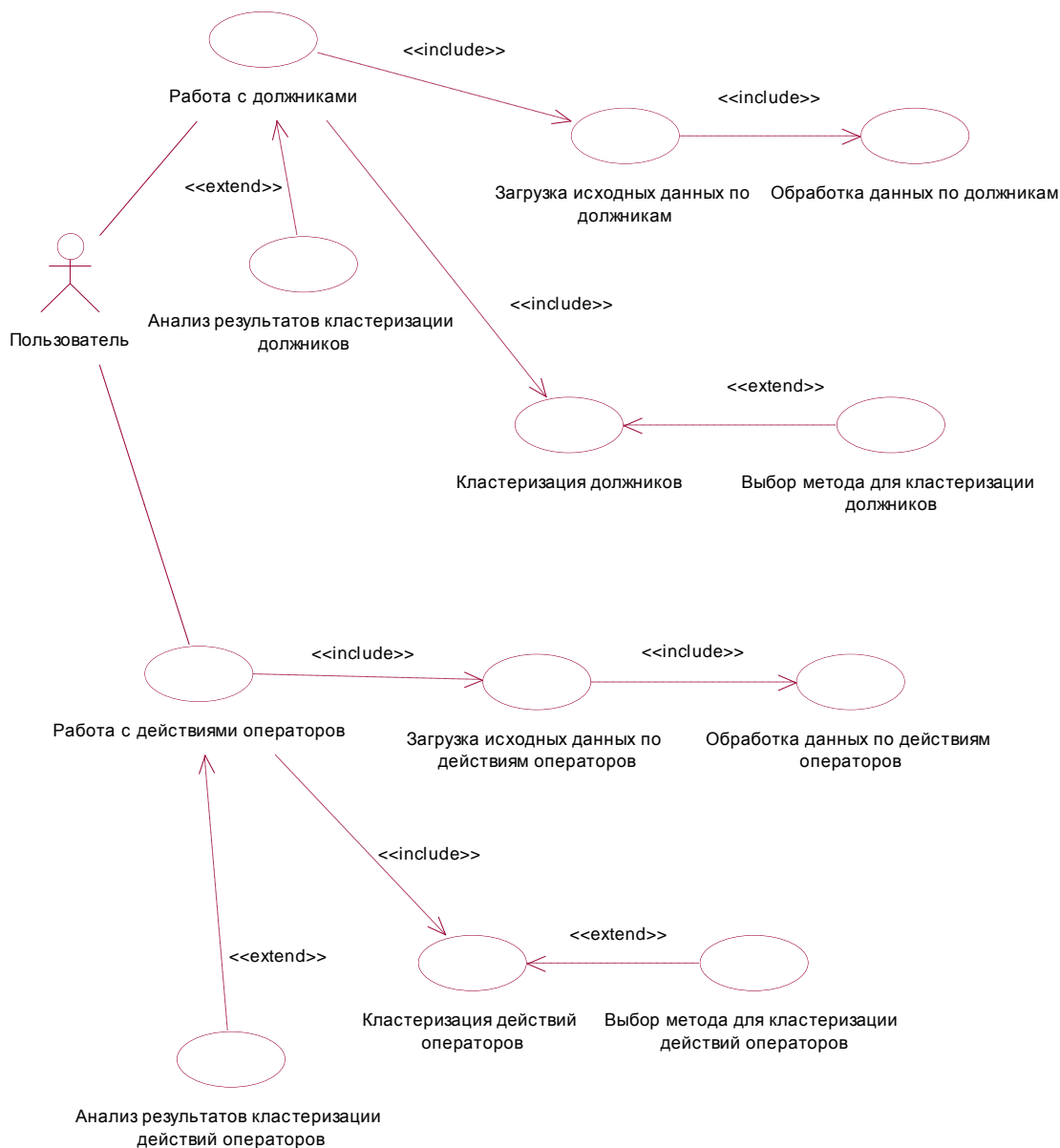


Рис. 1. Діаграма варіантів використання

Структуру системи в термінах об'єктно-орієнтованого проектування показує діаграма класів (рис. 2).

Користувач асоціативно пов'язаний з керуючим класом «Clustering» (головна програма). Класи «Дії операторів (таблиця)» й «Боржники (таблиця)» поєднані з керуючим класом відношенням агрегації, кожний з них асоціативно пов'язаний з файлами «Дії операторів (база даних)» й «Боржники (база даних)» відповідно.

Керуючий клас містить чотири атрибути – «Таблиці боржників», «Кластери боржників», «Таблиця дій операторів» й «Кластери дій операторів», які містять дані по боржниках і діям операторів відповідно (підготовлені для кластеризації й уже оброблені). Операції керуючого класу забезпечують підготовку даних, кластеризацію й подання результатів кластеризації (переклад умовних нормалізованих даних у колишній вид) для таблиць боржників і дій операторів.

Бази даних боржників і дій операторів – це файли, що виходять у результаті запиту до MySQL-сервера білінгової системи (операція «надати»), таблиці цих же даних виходять у результаті перетворення (операція «сформувані»).

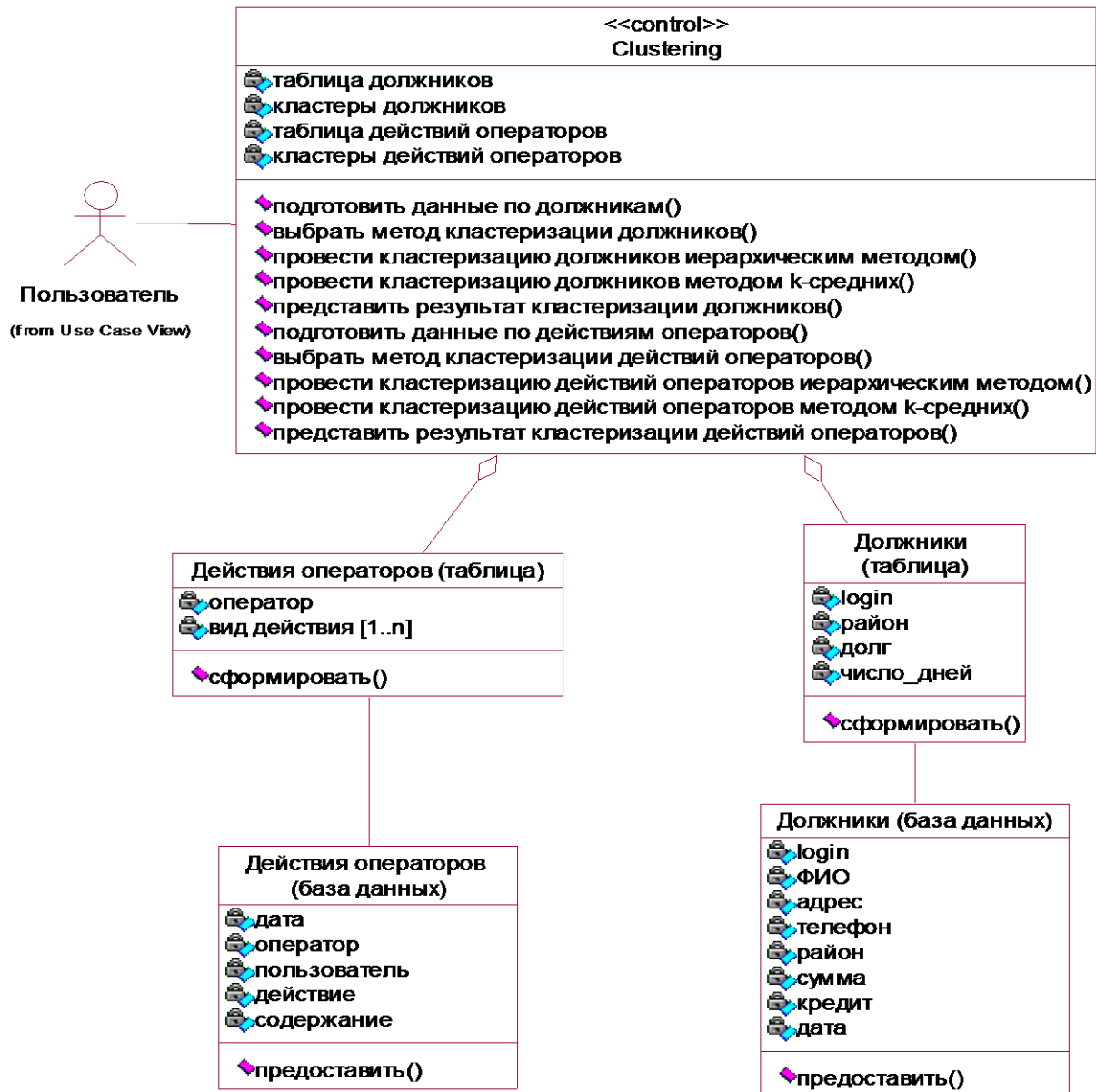


Рис. 2. Діаграма класів

На рис. 3 наведено діаграму станів типової роботи системи. Кожен зі станів «Підготовка даних по...» припускає низку внутрішніх дій: спочатку завантажується файл із даними, потім ці дані обробляються, і далі формуються таблиці.

Створена інформаційна модель системи для інтелектуального аналізу даних інтернет-провайдера була реалізована в середовищі Object Pascal. Інтерфейс застосування містить дві вкладки – «Боржники» й «Дії операторів». Кожна із вкладок оснащена трьома-чотирма кнопками («Завантажити дані», «Виконати розрахунок», «Показати в таблиці» й «Вивести дані»), таблицею для відображення вихідних і розрахункових даних, полем виводу для відображення ходу й результатів розрахунків, текстовим полем із приєднаним лічильником для вказівки номера кластера.

Завантажена інформація вже містить розрахований борг і число днів, що пройшли від дня останнього виходу в інтернет. Натискання кнопки «Виконати розрахунок» запускає алгоритм ієрархічної кластеризації, хід якого супроводжується коментарями в нижній частині вікна застосування. У це ж поле виводяться результати кластеризації (рис. 4).

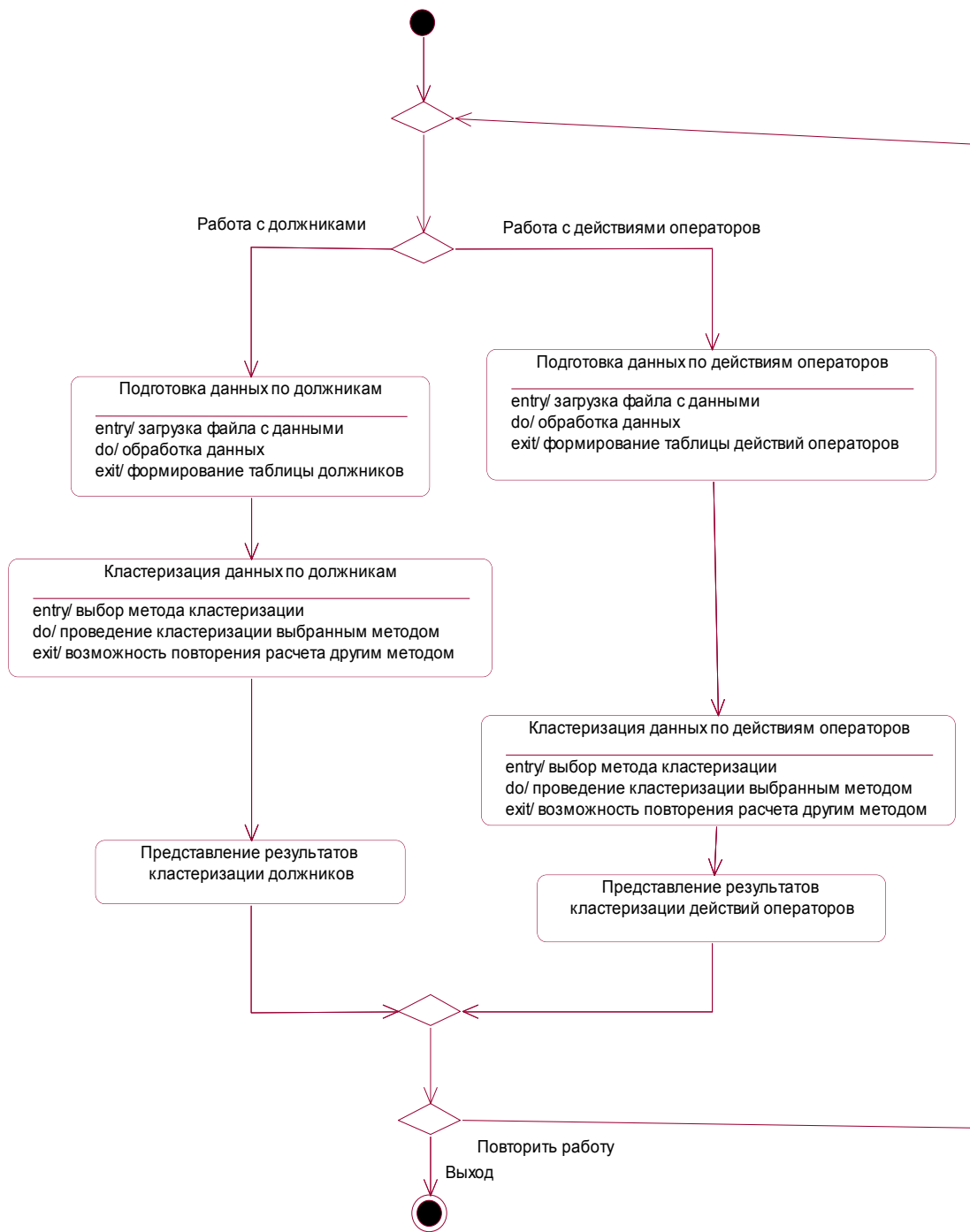


Рис. 3. Діаграма станів

Аналіз результатів кластеризації показує, що наявні в нас 18 боржників можуть бути згруповані в 7 груп (кластерів):

1. Абонент із самим більшим боргом, але дуже маленьким строком заборгованості (менш тижня) – вочевидь, людина не оплатила по якійсь поважній причині й у найскоріший час усуне проблему.

2. Абонент із відносно великою сумою боргу, але досить тривалим строком заборгованості (більше року) – вочевидь, з ним треба було працювати індивідуально, причому давно.

3. Абоненти із середньою сумою боргу (50-150 гривень) і з не дуже тривалим строком заборгованості (до півроку), але без визначення району – для ухвалення рішення тут потрібно задати питання відповідному операторові.

4. Абоненти із сумою боргу близько 100 гривень, але досить тривалим строком заборгованості (більше двох років) – очевидно, ці борги можна вважати безнадійними й списувати.

5. Абоненти з невеликою сумою боргу (до 100 гривень) і різними строками заборгованості, але із певним районом підключення – це сама більша групи, що має потребу в активній обробці.

6. Абонент із сумою боргу близько 100 гривень і дуже тривалим строком заборгованості (більше трьох років) – вочевидь, цей борг, як й в 4-й групі, можна вважати безнадійним і списати.

7. Абонент із маленькою сумою боргу (30 гривень) і річним строком заборгованості – тут можна рекомендувати нагадати абонентові про наявність боргу, однак краще також порахувати безнадійним і списати.

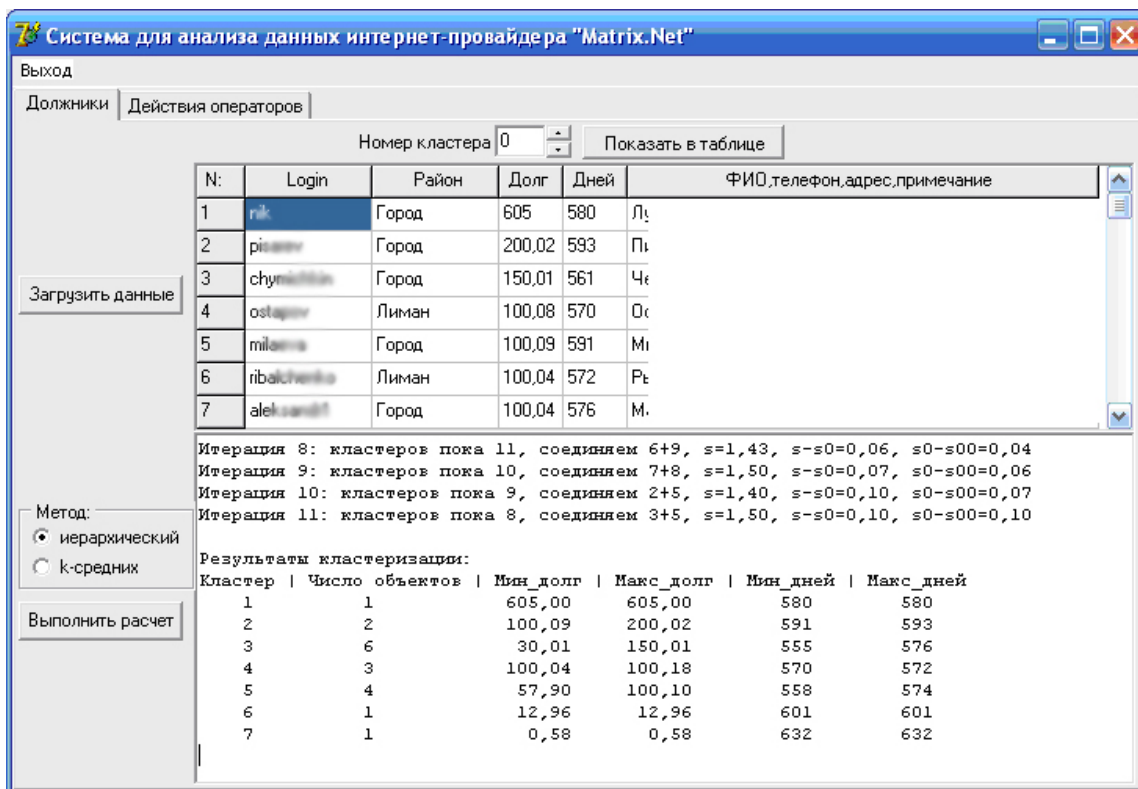


Рис. 4. Розрахунок кластеризації боржників

Якщо змінити метод розрахунку з ієрархічного, який було обрано за замовчуванням, на ітеративний метод к-середніх, та обрати розподіл на 3 кластера, можна побачити, що перший кластер містить 11 позицій, другий – 4, третій – 3. Але детальний аналіз отриманих груп (кластерів) свідчить, що ітеративний метод вважає найважливішим показником район міста. Зміна розподілу на сім кластерів замість трьох з першого погляду забезпечує результати, як і попередній метод кластеризації, але більш детальний аналіз цих результатів свідчить, що район міста також вважається найважливішим показником. Тому можна зробити висновок, що для розподілу боржників на групи краще підходить ієрархічний агломеративний метод, ніж ітеративний метод к-середніх.

Робота із вкладкою «Дії операторів» схожа на роботу з боржниками. Представлена в таблиці інформація вже містить розраховане число кожної дії кожного оператора. Натискання кнопки «Виконати розрахунок» запускає алгоритм ієрархічної кластеризації, хід якого супроводжується коментарями в нижній частині вікна застосунку. У це ж поле виводяться результати кластеризації (рис. 5).

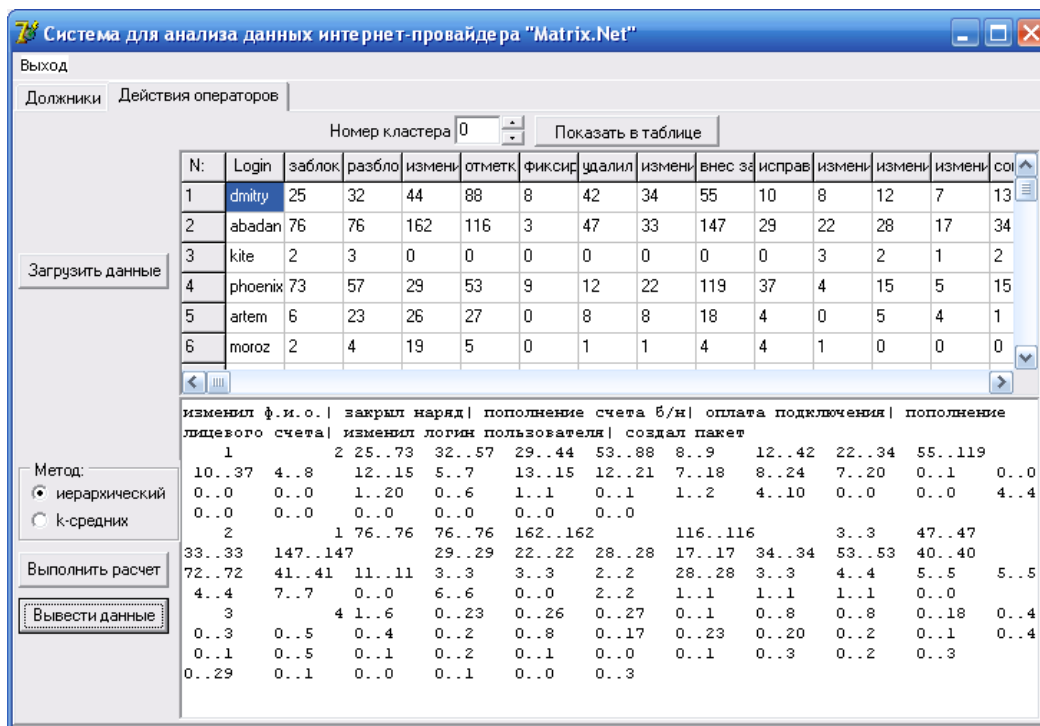


Рис. 5. Розрахунок кластеризації дій операторів

Уся інформація також може бути виведена в текстовий файл шляхом виклику контекстного меню, однак вона не дуже ясна для розуміння. Щоб представити результат розрахунку в більш наочному виді, необхідно натиснути кнопку «Вивести дані»: в XLS-таблицю будуть виведені значення числа дій по кожному операторі, що входить у кожен кластер, а також узагальнена інформація з кожного кластера (мінімум, максимум і середнє арифметичне). На рис. 6 наведено частину такої таблиці.

Як показано, сім операторів можуть бути згруповані в три групи (три кластери): максимальної активності, середньої активності і малої активності. Керівництву підприємства варто звернути увагу на першу (чим викликана така активність?) і третю (чим же займаються ці оператори?) групу.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA
№	Оператор	заблокировал пользователя	разблокировал пользователя	изменил кредит пользователю	отметка о выполнении задания	ф.и.о. наряд	удалил задание	изменил дату наряда	внес заявку	исправил заявку	изменил дату окончания пакета	изменил паспортные данные	изменил телефон	соглашение	добавил задание	создал пользователя	изменил пароль пользователю	сменил пароль пользователю	закрыл п/в	оплата аренды внешнего ip адреса	изменил ip	удалил платеж	изменил район пользователю	изменил тас пользователю	изменил адрес	за услуги клиента
Кластер номер 1																										
1	dmitry	25	32	44	88	8	42	34	55	10	8	12	7	13	21	18	24	20	0	0	0	0	20	0	1	0
2	phoenix	73	57	29	53	9	12	22	119	37	4	15	5	15	12	7	8	7	1	0	0	0	0	1	6	1
	Среднее	49	44,5	36,5	70,5	8,5	27	28	87	23,5	6	13,5	6	14	16,5	12,5	16	13,5	0,5	0	0	0	10,5	3	1	0,5
Кластер номер 2																										
1	abadan	76	76	162	116	3	47	33	147	29	22	28	17	34	53	40	72	41	11	3	3	2	28	3	4	5
Кластер номер 3																										
1	kite	2	3	0	0	0	0	0	0	0	3	2	1	2	0	1	5	1	2	1	4	1	1	0	1	0
2	moroz	2	4	19	5	0	1	1	4	4	1	0	0	0	2	2	10	2	2	0	0	0	4	1	1	0
3	tolik	1	0	0	9	1	3	6	1	0	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0
4	artem	6	23	26	27	0	8	8	18	4	0	5	4	1	8	17	23	20	2	0	0	0	5	1	2	1
	Минимум	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Максимум	6	23	26	27	1	8	8	18	4	3	5	4	2	8	17	23	20	2	1	4	1	5	1	2	1
	Среднее	2,75	7,5	11,3	10,3	0,25	3	3,75	5,75	2	1,25	2	1,25	1	2,5	5	9,5	5,75	1,5	0,25	1	0,25	2,5	0,5	1	0,25

Рис. 6. Результат кластеризації дій операторів: таблиця

Якщо змінити метод розрахунку з ієрархічного, який було обрано за замовчуванням, на ітеративний метод k-середніх з розподілом на 3 кластера, кластер максимальної активності містить чотири оператора, середньої активності – двох, малої активності – одного, що відрізняється від результатів ієрархічного методу. Але зміна числа кластерів з 3 на 2 приводить до прогнозованого результату – ми отримали той же кластер максимальної активності з чотирма операторами і кластер малої активності з трьома.

ВИСНОВКИ

Було проаналізовано низку методів інтелектуального аналізу даних, які можуть бути застосовані для обробки інформації інтернет-провайдера з метою удосконалення процесу прийняття управлінських рішень на підприємстві. Спроектовано інформаційну систему (застосування), яка дозволяє проводити розрахунки обраними методами, здійснено її програмну реалізацію.

Аналіз розрахунків дозволив зробити висновки:

- для аналізу даних щодо боржників та їх розподілу на групи краще підходить ієрархічний агломеративний метод, оскільки ітеративний метод k-середніх при кожному наступному використанні повертає різні результати;
- для аналізу даних щодо дій операторів та їх розподілу на групи відповідно цих дій краще підходить ієрархічний агломеративний метод, але також може бути використаний ітеративний метод k-середніх.

Розроблене застосування не тільки проводить інтелектуальний аналіз даних двома різними методами кластеризації, але здійснює попередню обробку даних з перетворенням їх до потрібного формату, що не може зробити жодне поширене стандартне застосування.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Білінгова система ABills. URL: <http://abills.net.ua/> (дата звернення 20.01.2020).
2. Carbon Billing 5. URL: <https://www.carbonsoft.ru/> (дата звернення 20.01.2020).
3. OneBox. URL: <https://crm-onebox.com/ru/onebox-this/> (дата звернення 20.01.2020).
4. Ubilling. URL: <http://ubilling.net.ua/> (дата звернення 20.01.2020).
5. Чубукова И.А. Data Mining: учебное пособие. Москва: Интернет-Университет Информационных Технологий; БИНОМ. Лаборатория знаний, 2006. 382 с.
6. Дюк В. А., Самойленко А. П.. Data Mining: учебный курс. Санкт-Петербург: Питер, 2001. 424 с.
7. Боровиков В.П. Искусство анализа данных на компьютере. Для профессионалов. Санкт-Петербург: Питер, 2001. 608 с.
8. Интернет провайдер MatrixNET. URL: <http://matrixnet.dn.ua/> (дата звернення 20.01.2020).
9. Деревя рішення. URL: <http://www.intuit.ru/studies/courses/6/6/lecture/174?page=1> (дата звернення 20.01.2020).
10. BaseGroup Labs: официальный сайт. URL: <https://basegroup.ru/deductor/function/algorithm/association-rules> (дата звернення 20.01.2020).
11. Штучна нейронна мережа. URL: https://uk.wikipedia.org/wiki/Штучна_нейронна_мережа (25.01.20).
12. Прогнозування за допомогою нейронних мереж. URL: https://wiki.tntu.edu.ua/Прогнозування_за_допомогою_нейронних_мереж (дата звернення 25.01.20).
13. Нейронні мережі – шлях до глибинного навчання. URL: <https://codeguida.com/post/739> (дата звернення 25.01.20).
14. Крынев А.В., Лукин Г. В. Математические методы обработки неопределенных данных. Москва: ФИЗМАТЛИТ, 2003. 216 с.
15. Мельников А. Ю. Объектно-ориентированный анализ и проектирование информационных систем: учебное пособие. Краматорск: ДГМА, 2013. 172 с.
16. Мельников О. Ю., Спориш Д. К. Об'єктне проектування системи для аналізу даних інтернет-провайдера провайдера. *Інформаційні технології в культурі, мистецтві, освіті, науці, економіці та бізнесі* : матеріали Міжнародної науково-практичної конференції. М-во освіти і науки України; М-во культури України; Київ. нац. ун-т культури і мистецтв. Ч. 1. Київ : Видавничий центр КНУКіМ, 2019. С. 59-60.

REFERENCES

1. Bilingov system ABills. URL: <http://abills.net.ua/> (date of the beast 01/20/2020).
2. Carbon Billing 5. URL: <https://www.carbonsoft.ru/> (date of burning 01/20/2020).

3. OneBox. URL: <https://crm-onebox.com/ru/onebox-this/> (date of the act 20.01.2020).
4. Ubilling. URL: <http://ubilling.net.ua/> (date of the beast 01/20/2020).
5. Chubukova I.A. Data Mining: textbook. Moscow: Internet University of Information Technologies; BINOMIAL. Knowledge Laboratory, 2006. 382 p.
6. Duke V. A., Samoylenko A. P. Data Mining: training course. St. Petersburg: Peter, 2001. -- 424 p.
7. Borovikov V.P. The art of data analysis on a computer. For professionals. St. Petersburg: Peter, 2001. 608 p.
8. Internet provider MatrixNET. URL: <http://matrixnet.dn.ua/> (date of the beast 01/20/2020).
9. Rishen tree. URL: <http://www.intuit.ru/studies/courses/6/6/lecture/174?page=1> (date of the beating 01/20/2020).
10. BaseGroup Labs: official site. URL: <https://basegroup.ru/deductor/function/algorithm/association-rules> (date of the blast 01/20/2020).
11. Piece of neural framing. URL: https://uk.wikipedia.org/wiki/Stuchna_neuronna_mereja (25.01.20).
12. Prediction behind the help of neural fissures. URL: https://wiki.tntu.edu.ua/Prediction_for_the_help_of_neural_freights (date of the act 25.01.20).
13. Neuronni frizzy - the way to go to bed. URL: <https://codeguida.com/post/739> (date of the beast is 01/25/2020).
14. Kryanev A.V., Lukin G.V. Mathematical methods of processing uncertain data. Moscow: FIZMATLIT, 2003. 216 p.
15. Melnikov A.Yu. Object-oriented analysis and design of information systems: a tutorial. Kramatorsk: DSMA, 2013. 172 p.
16. Melnikov O.Yu., Sporish D.K. About the design of the systems for the analysis of data from the Internet provider of the provider. *Information technologies in culture, mystery, education, science, economics and business: materials of the International Scientific and Practical Conference*. Ministry of Education and Science of Ukraine; Ministry of Culture of Ukraine; Kiev. nat. University of Culture i Mistress. Ch. 1. Kiev: Vidavnychy Center KNUKiM, 2019. P.59-60.

АВТОРИ / АВТОРЫ / AUTORS

Мельников О. Ю. – канд. техн. наук, доц. кафедри ІСПР ДДМА;
Мельников А. Ю. – канд. техн. наук, доц. кафедри ІСПР ДГМА;
Melnikov O. Yu. – candidate of technical science, associate professor DSEA.
E-mail: alexandr@melnikov.in.ua

Спориш Д. К. – магістр кафедри ІСПР ДДМА;
Спориш Д. К. – магист кафедры ИСПР ДГМА;
Sporysh D. K. – master DSEA.

Донбаська державна машинобудівна академія (ДДМА), м. Краматорськ.
Донбасская государственная машиностроительная академия (ДГМА), г. Краматорск.
Donbass State Engineering Academy (DSEA), Kramatorsk.

АНОТАЦІЯ / АННОТАЦИЯ / ANNOTATION

Мельников О. Ю., Спориш Д. К. Розробка системи для інтелектуального аналізу даних інтернет-провайдера. Вісник Донбаської державної машинобудівної академії. 2020. № 1 (48).

Метою роботи є удосконалення процесу прийняття управлінських рішень на підприємстві інтернет-провайдера шляхом дослідження методів інтелектуального аналізу даних та розробки інформаційної системи для проведення цього аналізу. Проаналізовано програмне забезпечення, призначене для обліку й тарифікації послуг, надаваних операторами зв'язку; здійснено опис і порівняння білінгових систем ABills, Carbon Billing 5, OneBox. Описана сутність інтелектуального аналізу даних, його задачі та методи, які використовуються для рішення цих задач. Розглянуто задачі класифікації та кластеризації, проаналізовано методи дерев рішень, штучних нейронних мереж, ієрархічних та ітеративних методів рішення задачі кластеризації. Застосування кожного методу проілюстровано рішенням відповідної задачі щодо даних інтернет-провайдера. Зроблено висновки про недоліки та переваги кожного методу. Наведено формалізацію задачі інтелектуального аналізу даних інтернет-провайдера, постановку та рішення задачі кластеризації з використанням ієрархічного методу та ітеративного методу k-середніх. Описано створену інформаційну модель спроектованої комп'ютерної системи мовою візуального моделювання UML. Описані можливості системи для інтелектуального аналізу даних, наведено приклад функціонування цієї системи та проведено аналіз результатів розрахунків. Оцінена ефективність розробки системи за допомогою функціонально-вартісного аналізу. Наукова новизна роботи полягає в застосуванні низки

методів для рішення задач інтелектуальної обробки конкретних наявних даних інтернет-провайдера. Практична цінність полягає в тому, що розроблено комп'ютерну систему, яка дозволяє проводити аналіз даних інтернет-провайдера для подальшого удосконалення процесу прийняття управлінських рішень на підприємстві.

Ключові слова: інтернет-провайдер, білінг, інтелектуальний аналіз даних, класифікація, кластеризація, метод дерев рішень, метод штучних нейронних мереж, ієрархічний агломеративний метод, метод k-середніх, інформаційна система, модель, unified modeling language.

Мельников А. Ю., Спорыш Д. К. Разработка системы для интеллектуального анализа данных интернет-провайдера. Вестник Донбасской государственной машиностроительной академии. 2020. № 1 (48).

Целью работы является усовершенствование процесса принятия управленческих решений на предприятии интернет-провайдера путем исследования методов интеллектуального анализа данных и разработки информационной системы для проведения этого анализа. Проанализировано программное обеспечение, предназначенное для учета и тарификации услуг, оказываемых операторами связи; осуществлено описание и сравнение нескольких биллинговых систем. Описана сущность интеллектуального анализа данных, его задачи и методы, которые используются для решения этих задач. Рассмотрены задача классификации и кластеризации, проанализированы методы деревьев решений, искусственных нейронных сетей, иерархических и итеративных методов решения задачи кластеризации. Применение каждого метода проиллюстрировано решением соответствующей задачи относительно данных интернет-провайдера. Сделаны выводы о недостатках и преимуществах каждого метода. Приведена формализация задачи интеллектуального анализа данных интернет-провайдера, постановка и решение задачи кластеризации с использованием иерархического метода и итеративного метода k-средних. Описана созданная информационная модель спроектированной компьютерной системы на языке визуального моделирования UML. Описаны возможности системы для интеллектуального анализа данных, приведен пример функционирования этой системы и проведен анализ результатов расчетов. Оценена эффективность разработки системы с помощью функционально-стоимостного анализа. Научная новизна работы заключается в применении ряда методов для решения задач интеллектуальной обработки конкретных имеющихся данных интернет-провайдера. Практическая ценность заключается в том, что разработана компьютерная система, которая разрешает проводить анализ данных интернет-провайдера для дальнейшего усовершенствования процесса принятия управленческих решений на предприятии.

Ключевые слова: интернет-провайдер, биллинг, интеллектуальный анализ данных, классификация, кластеризация, метод деревьев решений, метод искусственных нейронных сетей, иерархический агломеративный метод, метод k-средних, информационная система, модель, unified modeling language.

Melnykov O. Yu., Sporysh D. K. Development of the system for data mining of Internet service. Herald of the DSEA. 2020. № 1 (48).

The aim of this work is to improve the management decision-making process at the Internet Service Provider enterprise by researching data mining methods and developing an information system for this analysis. At first analyzes software designed for accounting and charging services provided by telecom operators; description and comparison of billing systems ABill, Carbon Billing 5, OneBox. Next describes the essence of data mining, its tasks and methods that are used to solve these problems. The classification and clustering problem is considered, methods of decision trees, artificial neural networks, hierarchical and iterative methods for solving the clustering problem are analyzed. The application of each method is illustrated by solving the corresponding problem regarding the data of the Internet provider. Conclusions are made about the disadvantages and advantages of each method. The third section of presents the formalization of the data mining problem of the Internet provider, the statement and solution of the clustering problem using the hierarchical method and the iterative k-means method. The created information model of the designed computer system in the language of visual modeling UML is described. The fourth section describes the capabilities of the system for data mining, provides an example of the functioning of this system and analyzes the calculation results. The efficiency of the development of the system using the functional-cost analysis is estimated. The scientific novelty consists in the application of a number of methods to solve the problems of intellectual processing of specific available data from an Internet provider. The practical value lies in the fact that a computer system has been developed that allows you to analyze the data of the Internet provider to further improve the management decision-making process at the enterprise.

Keywords: internet service providers, billing, data mining, classification, clustering, simulation, decision trees, the method of artificial neural networks, hierarchical agglomerative method, k-means, information system, models, unified modeling language.