

РЕФЕРАТ

Волошина Т.Н. Интеллектуальное пространство украинского литературного языка / Татьяна Николаевна Волошина // Управление проектами, системный анализ и логистика. Вестник НТУ. — К.: НТУ. — 2012. — Вып. 26.

В статье исследовано интеллектуальные измерения украинского литературного языка в контексте социокультурного прогресса этноса, функциональных возможностей языка в коммуникативных типах речи, тенденции развития современного украинского литературного языка.

Актуальность проблемы интеллектуального развития языка определяется пониманием вопросов измерения мышления и языка, цивилизационной эволюции языковых единиц как форм динамического сознания, анализа языковых ресурсов и их возможностей в контексте социокультурного прогресса этноса, функциональных возможностей языка в коммуникативных типах речи, тенденций развития современного украинского литературного языка.

Современные лингвистические исследования сориентированы на аргументации языка в его познавательных и цивилизационных параметрах, а основное внимание сконцентрировано на языке как мыслительной форме, как сущности, как функции и трансформации и как процессе развития. Лингвистическая проблематика пересекается с эволюционными измерениями бытия, что определяет основные тенденции аспектологии современного теоретического языкознания, соответственно интеллектуализация языка сконцентрирована на проблеме объединения тенденций синтезировать лингвистические знания с философией, психологией, эстетикой, историей, комплексом проблем аргументации интеллектуальных возможностей языка, проблем функциональной стилистики и мотивации языкового прогресса.

Исследовательская проблематика интеллектуальных возможностей языка конкретизуется в лингвофилософских, лингволингвистических и лингвопсихологических аспектах: на общих принципах предметной реализации естественного языка; понятиях “мира” и “ситуации” и их языковых формах как отдельного научного языка; особенностях восприятия и факторах знаний, представлений о действительности в формировании высказывания. Лингвистическая традиция анализа языковой картины мира как интеллектуальной репрезентации познавательных процессов мышления связана прежде всего с теорией внутренней формы слова А.А. Потебни и далее в работах украинских лингвистов: Л.А. Булаховского, В.М. Русановского, Т.Р. Кияка, С.Я. Ермоленко, Т.И. Панько, Л.О. Пустовит, В.С. Перебийноса, М.П. Кочергана, М.Я. Плющ.

КЛЮЧЕВЫЕ СЛОВА: ИНТЕЛЛЕКТУАЛЬНОЕ ПРОСТРАНСТВО ЯЗЫКА, ИЗМЕРЕНИЯ МЫШЛЕНИЯ И ЯЗЫКА, ЦИВИЛИЗАЦИОННАЯ ЭВОЛЮЦИЯ ЯЗЫКОВЫХ ЕДИНИЦ, ФУНКЦИОНАЛЬНЫЕ ВОЗМОЖНОСТИ ЯЗЫКА, КОММУНИКАТИВНЫЕ ТИПЫ РЕЧИ.

УДК 004.45

МЕТОДИ ТА ТЕХНОЛОГІЇ ІНТЕЛЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ

Гавриленко В.В., доктор фізико-математичних наук
Іванченко Г.Ф., кандидат технічних наук
Шевченко Г.Є.

Актуальність дослідження. Для аналізу даних і отримання нових знань використовують методи Data Mining. У науковій літературі для позначення напряму інтелектуального аналізу найчастіше використовуються терміни «Data Mining» і «Knowledge Discovery», в україномовних публікаціях з даної тематики застосовується термін "дейтамайнінг" (надалі – ДМ) або «інтелектуальний аналіз даних» (надалі - ІАД). Порівняно недавно на перехресті декількох областей інформаційних систем і технологій сформувався і активно розвивається цей науковий напрям, метою якого є пошук корисних знань в корпоративних базах і сховищах даних.

Постановка проблеми. Одним з актуальних завдань для сучасних спеціалістів в області інформаційних технологій є наступні: володіння технологіями дейтаманінгу, розуміння його можливостей, вміння використовувати основні операції зі створення та ефективного застосування відповідних програмних засобів.

Теоретичний аналіз дослідження. Останнім часом для вирішення практичних завдань все частіше застосовуються методи інтелектуального аналізу даних (Data Mining). ІАД — виявлення

прихованих закономірностей або взаємозв'язків між змінними у великих масивах необроблених даних. ІАД є досить новою галуззю інформаційних технологій, яка постійно розвивається. Бурхливий розвиток програмних систем ДМ привів до необхідності створення деяких стандартів у цій сфері. Через відсутність міжнародних стандартів, стали з'являтися стандарти окремих фірм-виробників, корпорацій і окремих експертів, що намагаються упорядкувати і погодити досягнення всієї індустрії ДМ за останнє десятиліття.

Стандарти повинні торкатися трьох основних аспектів технології дейтамайнінгу: *по-перше*, уніфікації інтерфейсів, за допомогою яких будь-яка прикладна програма може одержати доступ до функціональності ДМ. Тут склалося два напрямки. Це стандартизація інтерфейсів для об'єктних мов програмування (CWM Data Mining, JDM) і спроби розробки надбудови над мовою SQL, що дозволяла б звертатися до інструментарію ДМ убудованого безпосередньо в реляційну базу даних (Microsoft OLE DB for Data Mining, SQL/MM, DMQL, MSOL, MINE RULE); *по-друге*, це вироблення єдиної угоди по збереженню і передачі моделей Data Mining. Основою для подібного стандарту є мова XML. Сам стандарт зветься PMML (Predicted Model Markup Language); *третьій* аспект представляє стандарт CRISP, що дає рекомендації з організації процесу Data Mining у цілому.

Як підсумок аналізу першої підгрупи стандартів, слід зазначити, вони не вирішують остаточно проблему взаємодії між ДМ-пакетами та DB у великих БД. Застосовуючи ДМ у великих БД постає потреба у тому, щоб ці мови/стандарти ефективно виконували дві задачі: селективна генерація шаблонів та управління попередньо згенерованими результатами (в т.ч. і під час минулих сесій ДМ). Отже, описані SQL-орієнтовані мови ДМ в подальшому повинні розвиватися в напрямку здатності створення окремого стійкого та надійного сховища результатів сесій ДМ у надвеликих базах даних.

PMML, як представник другої групи стандартів, забезпечує швидкий і простий спосіб визначення статистичних моделей і ДМ-моделей, і можливість вільного обміну і спільного використання цих моделей програмними продуктами різних виробників. PMML є загальним, незалежним від розробника і конкретної реалізації продукту ДМ некомерційним інтерфейсом, що дає можливість програмам різних виробників обмінюватися отриманими моделями і закономірностями, не розкриваючи при цьому комерційної власності кожного з них. Це дозволяє користувачам розробляти моделі для додатків одного виробника, і використовувати додатки інших виробників для аналізу, оцінки і використання моделі.

CRISP-DM (The Cross Industrie Standard Process for Data Mining) - це міжгалузевий стандарт процесу аналізу даних і підтримки прийняття рішень. Він з'явився в наслідок роботи міжнародного консорціуму в складі: *NCR Systems Engineering Copenhagen* (Данія), *DaimlerChrysler AG* (Німеччина), *Integral Solutions Ltd.* (Англія) і *OHRA Verzekeringen en Bank Groep B.V.* (Голландія). Цей проект також підтримується Європейською комісією в рамках програми ESPRIT. Наразі для вільного використання та обговорення доступна версія 1.0 стандарту.

CRISP-DM проект створює процесно-орієнтовану ДМ-модель, що є нейтральною по відношенню до програмних засобів ДМ та до галузі застосування. Починаючи з мало розвинутих на сьогодні процесів ДМ у народному господарстві, та точно відповідаючи потребам кінцевих користувачів, CRISP-DM проект визначає та оцінює ДМ-процес, який можна застосовувати в різних секторах народного господарства. Це прискорює великі ДМ-проекти, робить їх більш дешевими, більш достовірними та керованими. Навіть малі ДМ проекти вииграють від використання CRISP.

Модель процесу ДМ у цьому стандарті – це опис життєвого циклу проекту ДМ у будь-якій прикладній галузі. Ця модель містить фази проекту, задачі, що відповідають їм, і зв'язки між цими задачами. Послідовність фаз закріплена не жорстко. Допускається переміщення назад і вперед між різними фазами, що відповідає ітеративному характеру процесів розробки і необхідності іноді по-новому глянути на проблему при одержанні проміжних результатів. Відповідно до цього стандарту кожний проект ДМ, поза залежністю від галузі додатка, повинен містити наступні фази: – Розуміння бізнес-проблеми; – Розуміння даних; – Підготовка даних; – Побудова моделі; – Обчислення; – Реалізація результату.

Можна виділити п'ять загальних типів закономірностей (паттернів), що можуть бути одержані засобами дейтамайнінгу:

- *Класифікація*: робить висновок щодо визначення характеристик конкретної групи (наприклад, споживачі, які були втрачені через дії конкурентів).

- *Кластеризація (групування)*: ототожнює групи елементів, які використовують спільно зображуючий параметр сигналу даних (кластеризація відрізняється від класифікації, бо наперед визначені характеристики не приводиться)

- *Асоціація*: ідентифікує зв'язки або відношення між подіями, які відбувалися колись (наприклад, зміст кошика відвідання магазину за покупкою).

- *Послідовність*: подібна до асоціації, крім того зв'язок ідентифікується за певні періоди часу (наприклад, повторний візит до супермаркету або фінансове планування продукту)
- *Прогнозування*: оцінює майбутні значення, основані на взірцях всередині великої набору даних (наприклад, прогнозування попиту).

З точки зору орієнтації на процес є три класи процесів ІАД: відкриття (discovery, добування); пророче моделювання (predictive modeling, моделювання передбачень); аналіз аномалій (forensic analysis).

Викладення основного матеріалу дослідження. Для набуття практичних навичок розв'язання задач класифікації можна залучити наступні програмні засоби ДМ: SPSS (SPSS, Clementine), Statistica (StatSoft), SAS Institute (SAS Enterprise Miner). Є постачальники, які включають Data Mining рішення в функціональність СУБД: це Microsoft (Microsoft SQL Server), Oracle, IBM (IBM Intelligent Miner for Data). Привертає увагу своїми перевагами комп'ютерна система ДМ бібліотеки Xelopes.

Xelopes – вільно поширювана бібліотека, що забезпечує універсальну основу для стандартного доступу до алгоритмів data mining. Вона була розроблена німецькою компанією ProdSys в тісній співпраці з фахівцями російської фірми ZSoft.

Для порівняння можна розглянути ціни на представлені вище продукти для бізнес-користувачів, що згруповані наступним чином:

- Понад 10000\$: IBM, Oracle, SAS, SPSS (рівень підприємства).
- Від 1000\$ до 10000\$: Microsoft SQL Server, Statsoft Statistica, ThinkAnalytics) (рівень відділу).
- До 1000\$: Excel, See5 (рівень користувача).
- Вільно поширюване програмне забезпечення: Xelopes.

Унікальність Xelopes полягає в його гнучкій архітектурі, що складається з трьох головних частин: потужної і розширюваної бібліотеки алгоритмів Data Mining, уніфікованого, заснованого на відкритих стандартах, інтерфейсу прикладного програмування, а також сумісність з такими стандартами data mining, як CWM і PMML.

Відповідно до стандарту CWM загальну концепцію роботи алгоритмів бібліотеки Xelopes можна представити так, як це зображено на рис. 1.

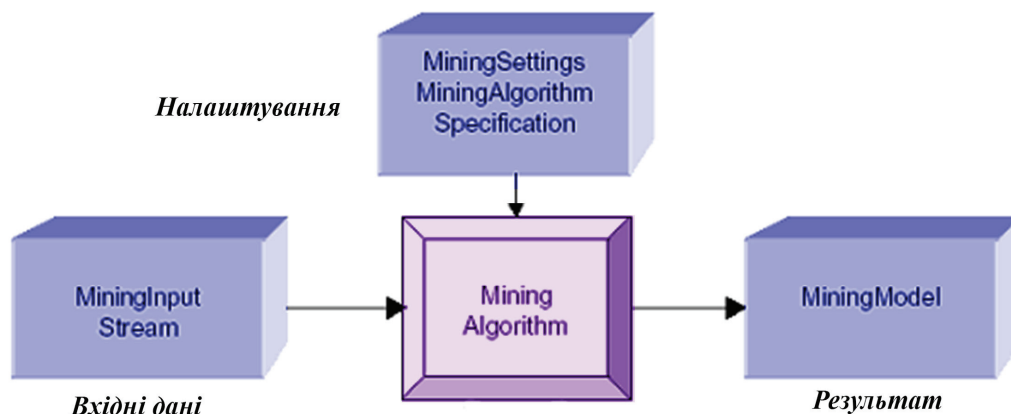


Рисунок 1. – Загальна концепція роботи алгоритмів в бібліотеці Xelopes

Згідно даній концепції на вхід будь-якого алгоритму подаються початкові дані у вигляді потоку, представленого класом *MiningInputStream*. Окрім даних перед роботою алгоритму повинні бути виконані відповідні настройки. Вони діляться на два типи: настройки, що відображають специфіку для вирішуваної задачі (моделі, що будується), і настройки, які специфічні для конкретного алгоритму. Перші реалізуються за допомогою спадкоємства від класу *MiningSettings*, другий тип – за допомогою спадкоємства від класу *MiningAlgorithmSpecification*. Результат роботи алгоритму представляються у вигляді моделей, які є екземплярами класу *MiningModel*. Самі алгоритми успадковуються від класу *MiningAlgorithm*.

У бібліотеці Xelopes реалізовані класи настройок, моделей і алгоритмів для основних задач data mining (табл. 1). Необхідно помітити, що вхідні дані не залежать від вирішуваної задачі і для будь-якого алгоритму представляються у вигляді екземпляра класу *MiningInputStream*.

Таблиця 1. – Основні класи бібліотеки Xelopes

№	Задача	Клас настроюк	Клас моделі	Клас алгоритму
1	Статистичні задачі	StatisticsMiningSettings	StatisticsMiningModel	StatisticsAlgorithm
2	Пошук асоціативних правил	AssociationRulesSettings	AssociationRulesMiningModel	AssociationRulesAlgorithm
3	Сіквінціальний аналіз	SequentialSettings	SequentialMiningModel	SequentialAlgorithm
4		CustomerSequentialSettings	CustomerSequentialMiningModel	CustomerSequentialAlgorithm
5	Кластеризація	ClusteringMiningSettings	ClusteringMiningModel	ClusteringAlgorithm
6	Регресії	SupportVectorSettings	SupportVectorMiningModel	SupportVectorAlgorithm
7		SparseGridsSettings	SparseGridsMiningModel	SparseGridsAlgorithm
8	Класифікації	DecisionTreeSettings	DecisionTreeMiningModel	DecisionTreeMiningModel

Для реалізації власного алгоритму необхідно реалізувати клас, успадкований від одного з підкласів класу *MiningAlgorithm*. В підкласі слід реалізувати метод *runAlgorithm()*. Даний метод викликається з методу вже реалізованого *buildModel()*. В цьому ж методі після запуску алгоритму здійснюється створення моделі відповідного підкласу класу *MiningModel*.

Алгоритм повинен враховувати настройки. Для цього в класі *MiningAlgorithm* реалізовані методи *setMiningSettings (MiningSettings miningSettings)* і *setApplicationInputSpecification (ApplicationInputSpecification applicationInputSpecification)*. Вхідні дані передаються в алгоритм за допомогою реалізованого в цьому ж класі методу *setMiningInputStream (MiningInputStream miningInputStream)*.

Таким чином, для додавання в бібліотеку Xelopes нового алгоритму достатньо реалізувати абстрактні методи одного з підкласів класу *MiningAlgorithm*. Особливості реалізації залежать від задачі, вирішуваної створюваним алгоритмом.

Системи підтримки ухвалення рішень використовують методи ДМ для аналізу даних і отримання нових знань. Одержані результати можуть носити як описовий характер, дозволяючи краще зрозуміти дані, так і передбачуваний, дозволяючи надалі передбачувати значення деяких параметрів на підставі знайдених закономірностей. До першого вигляду відносяться методи, що виконують пошук асоціативних правил і кластеризацію, до другого – методи, що будують функції класифікації і регресії.

Бібліотека Xelopes в своєму складі має алгоритми обох типів. Вона може служити основою для побудови СППР. Структура і підтримка бібліотекою основних стандартів роблять її використання достатньо простим.

Основний підхід рішення задач data mining за допомогою бібліотеки Xelopes не залежить від виду або методу, що використовується. Спочатку створюється екземпляр класу *MiningInputStream* для завантаження початкових даних, наприклад, з файлу формату ARFF:

```
MiningInputStream inputData = new MiningArffStream("data.arff");
```

Потім виділяються метадані завантажених даних:

```
MiningDataSpecification metaData = inputData.getMetaData();
```

Далі створюється екземпляр класу *MiningAlgorithm*. Для настройки процесу побудови моделі створюються екземпляри класів *MiningSettings* і *MiningAlgorithmSpecification*. У даних екземплярів встановлюються необхідні параметри. Створення конкретних екземплярів класів залежать від методу, що використовується, і вирішуваної задачі.

Будь-яка побудована модель може бути збережена у форматі PMML.

GUI інтерфейс бібліотеки ДМ алгоритмів дозволяє виконувати наступні основні функції:

- Завантажити дані, що представлені у вигляді текстового файлу формату arff, і проглянути їх в табличному вигляді;
- Одержати інформацію про атрибути даних (полях таблиці);
- Одержати статичну інформацію про початкові дані;
- Побудувати модель data mining;
- Візуалізувати побудовану модель (для асоціативних правил, дерев рішень і дейтограмм);
- Зберегти модель і застосувати її надалі.

У Xelopes розрізняють два основні типи атрибутів: категоріальний і числовий. Залежно від типу змінюється і інформація про атрибут. Для будь-якого атрибуту виводиться його назва і тип.

Для категоріальних атрибутів відображається інформація про значення, що приймаються: кількості (Number categories) і списку значень (Categories). Якщо кількість категорій не обмежено, то з'явиться прохання поставити прапорець на *unbounded categories*.

Для числових атрибутів відображається інформація про найбільше (Upper) і найменше (Lower) значення. Крім того, залежно від властивостей атрибуту можуть бути встановлені наступні прапорці: *Cyclic* – якщо значення атрибуту циклічні (тобто може бути визначено поняття відстані); *Discrete* – якщо значеннями атрибуту є дискретні величини; *Time* – якщо атрибут є часом.

У результаті використання методів data mining повинна бути побудована mining модель. Для цього необхідно натиснути кнопку Build Mining Model на панелі інструментів або вибрати пункт меню Model | Build. В результаті відкриється діалогове вікно, яке пропонує побудувати моделі для завантажених раніше даних (рис. 2).

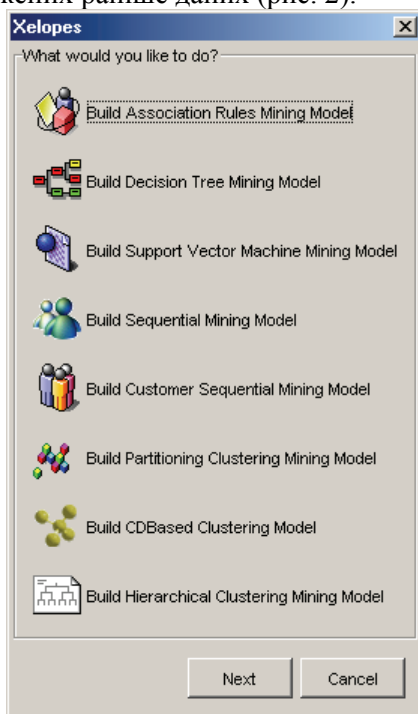


Рисунок 2. – Типи моделей, створюваних алгоритмами бібліотеки Xelopes.

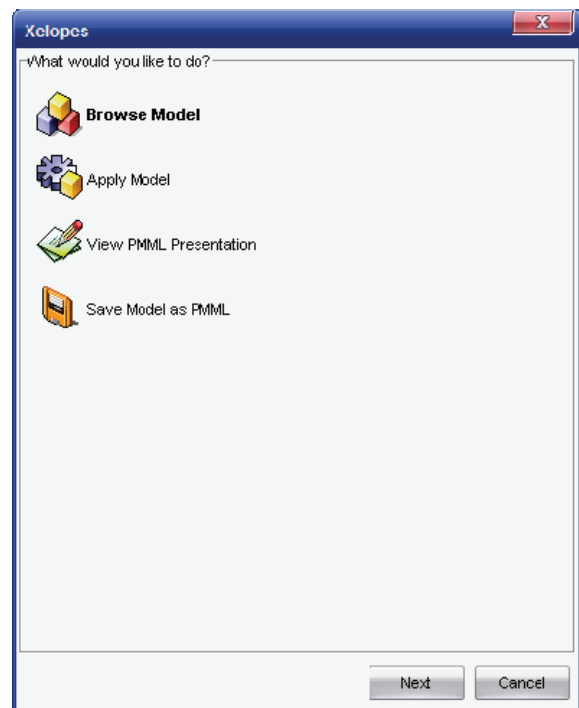


Рисунок 3. – Дії, що виконуються з побудованою моделлю.

Маємо для побудови наступні доступні моделі:

- асоціативні правила (Association Rules Mining Model);
- дерева рішень (Decision Tree Mining Model);
- математична залежність, що побудована методом SVM (Support Vector Machine Mining Model);
- послідовності (Sequential Mining Model);
- модель сиквенціального аналізу (Customer Sequential Mining Model);
- кластерна модель, що розділяється (Partition Clustering Mining Model);
- центрована кластерна модель (CDBased Clustering Mining Model);
- ієрархічна кластерна модель (Hierarchical Clustering Mining Model).

Після завершення побудови моделі з'явиться діалогове вікно, що пропонує виконати наступні дії (рис. 3):

- Візуалізувати модель (Browse Model).
- Застосувати модель (Apply Model).
- Показати модель у вигляді PMML (View PMML Presentation).
- Записати модель в PMML форматі (Save Model as PMML).

Приклад візуалізації асоціативних правил можна побачити на рис. 4.

Процес підготовки до ІАД припускає збір даних для аналізу з різних джерел даних і представлення їх у форматі, придатному для використання алгоритмів data mining.

Бібліотека Xelopes підтримує ARFF (Attribute-Relation File Format) формат представлення даних. Він розроблений для бібліотеки Weka в університеті Waikato. ARFF-файл є ASCII текстовим файлом, що описує список об'єктів із загальними атрибутами.

Структурно такий файл розділяється на дві частини: заголовок і дані.

Заголовок містить інформацію про ім'я файлу і метадані про представлені в ньому дані. Ім'я описується в наступному форматі

@relation <ім'я>

Метадані описують атрибути представлених у файлі даних. Інформація про кожний атрибут записується в окремому рядку і включає ім'я атрибуту і його тип. Очевидно, що імена повинні бути унікальні. Порядок їх опису повинен співпадати з порядком колонок опису даних. Загальний формат опису атрибуту наступний:

@attribute <ім'я атрибуту> <тип атрибуту>

Значенням поля <тип> може бути одне з наступних п'яти типів: real, integer, <категорія>, string, date [<формат дати>].

Типи real і integer є числовими. Категоріальні типи описуються переліком категорій (можливих значень). Наприклад: *@attribute outlook {sunny, overcast, rainy}*.

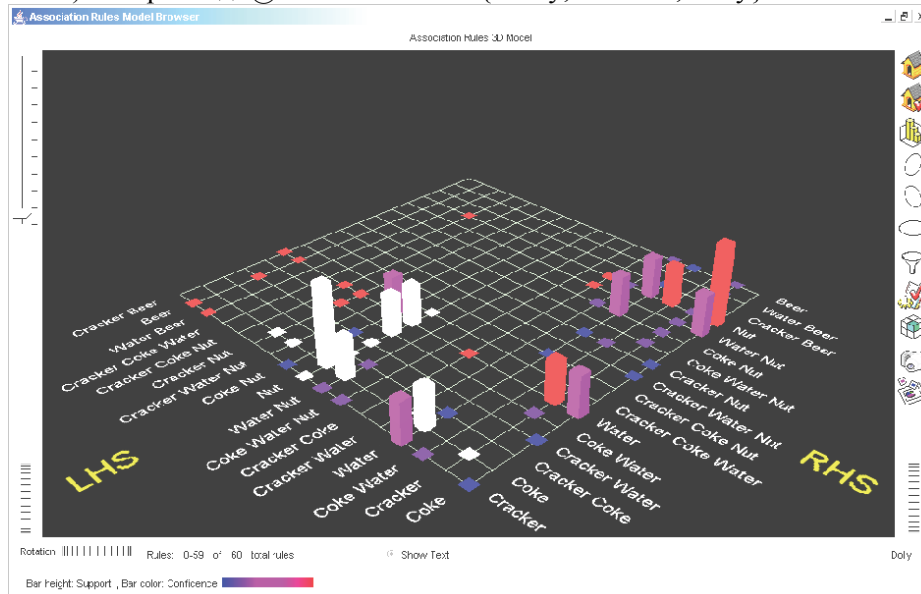


Рисунок 4. – Візуальне представлення асоціативних правил

Дані представляються в ARFF-форматі у вигляді списку значень атрибутів об'єктів після тега @data. Кожний рядок списку відповідає одному об'єкту. Кожний стовпчик відповідає атрибуту, описаному в частині заголовка. Причому порядок проходження стовпчиків повинен співпадати з порядком опису атрибутів.

Висновки. Розглянута вільно поширювана бібліотека алгоритмів Xelopes містить всі необхідні засоби для розв'язання задач аналізу ринкових кошиків, сіквенціального аналізу, кластерного аналізу, проведення класифікації методами decision tree і support vector machine. В даний час ведуться роботи по додаванню в бібліотеку нових алгоритмів. Вбудована в Xelopes підтримка мови PMML дозволяє їй обмінюватися інформацією з цілою низкою комерційних Data Mining продуктів. Завдяки своїй гнучкості і функціональності Xelopes є оптимальним вибором для побудови інформаційно-аналітичних систем в бізнесі, науці та освіті.

ПЕРЕЛІК ПОСИЛАНЬ

1. Барсегян А.А. Методы и модели анализа данных: OLAP и Data Mining / Барсегян А.А., Куприянов М.С., Степаненко В.В., Холод И.И. – СПб.: БХВ-Петербург, 2004. – 336 с.
2. Барсегян А.А. Технологии анализа данных: Data Mining, Visual Mining, Text Mining, OLAP / Барсегян А.А., Куприянов М.С., Степаненко В.В., Холод И.И. – СПб.: БХВ-Петербург, 2010. – 384 с.
3. Дюк В. Data Mining: учебный курс / Дюк В., Самойленко А. – СПб: Питер, 2001. – 368 с.

РЕФЕРАТ

Гавриленко В.В., Іванченко Г.Ф., Шевченко Г.Є. Методи та технології інтелектуального аналізу даних./ Валерій Володимирович Гавриленко, Геннадій Федорович Іванченко, Галина Єгорівна Шевченко // Вісник НТУ. — К.: НТУ. – 2012. – Вип. 26.

Стаття присвячена питанням впровадження на кафедрі інформаційних систем і технологій Національного транспортного університету нової нормативної дисципліни “Інтелектуальний аналіз

даних” для студентів, що навчаються за напрямом “Комп’ютерні науки”, та отриманому при цьому досвіді.

Об’єкт дослідження – комп’ютерна система дейтамайнінгу бібліотеки Xelopes

Мета роботи – визначення доцільності використання комп’ютерної системи дейтамайнінгу бібліотеки Xelopes.

Розглянута в статті бібліотека алгоритмів інтелектуального аналізу даних від компанії Prudsys містить всі необхідні засоби для розв’язання задач аналізу та дозволяє виконувати наступні основні функції: завантажити дані представлені у вигляді текстового файлу формату arff і проглянути їх в табличному вигляді; одержати інформацію про атрибути даних (полях таблиці); одержати статичну інформацію про початкові дані; побудувати модель data mining; зберегти модель і застосувати її надалі.

КЛЮЧОВІ СЛОВА: ІНТЕЛЕКТУАЛЬНИЙ АНАЛІЗ ДАНИХ, ДЕЙТАМАІНІНГ, БІБЛІОТЕКА XELOPES.

ABSTRACT

Gavrilenko V.V. Ivanchenko G.F., Shevchenko G.E. Methods and technologies for data mining. / Valery Gavrilenko, Gennady Ivanchenko, Galina Shevchenko // Visnik NTU. — К.: NTU . – 2012. – Vol. 26.

The article deals with the introduction of the Department of Information Systems and Technologies of the National Transport University new regulatory discipline "Data mining" for students in the field of "computer science", and of the resulting experience.

Object of study - computer system deytamayninga library Xelopes

Purpose - to determine the feasibility of using the library computer system deytamayninga Xelopes.

Discussed in the article library data mining algorithms from the company Prudsys contains all the necessary tools to address the problems of analysis and enables the following functions: load data is in a text file format arff and view them in tabular form; obtain information about the attributes of the data (table fields) get static information about the source data, build a model of data mining; save the model and apply it in the future.

KEY WORDS: DATA MINING, DEYTAMANING, LIBRARY XELOPES.

РЕФЕРАТ

Гавриленко В.В., Иванченко Г.Ф., Шевченко Г.Е. Методы и технологии интеллектуального анализа данных. / Валерий Владимирович Гавриленко, Геннадий Федорович Иванченко, Галина Егоровна Шевченко // Вестник НТУ. — К.: НТУ. – 2012. – Вып. 26.

Статья посвящена вопросам внедрения на кафедре информационных систем и технологий Национального транспортного университета новой нормативной дисциплины "Интеллектуальный анализ данных" для студентов, обучающихся по направлению "Компьютерные науки", и полученном при этом опыте.

Объект исследования – компьютерная система дейтамайнинга библиотеки Xelopes

Цель работы – определение целесообразности использования компьютерной системы дейтамайнинга библиотеки Xelopes.

Рассмотренная в статье библиотека алгоритмов интеллектуального анализа данных от компании Prudsys содержит все необходимые средства для решения задач анализа и позволяет выполнять следующие основные функции: загрузить данные представлены в виде текстового файла формата arff и просмотреть их в табличном виде; получить информацию об атрибутах данных (полях таблицы); получить статическую информацию о исходные данные; построить модель data mining; сохранить модель и применить ее в дальнейшем.

КЛЮЧЕВЫЕ СЛОВА: ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ДАННЫХ, ДЕЙТАМАІНІНГ, БІБЛІОТЕКА XELOPES.