УДК 681.2

СПОСОБИ УРАХУВАННЯ НЕВИЗНАЧЕНОСТІ ПРИ ПОБУДОВІ ЛІНГВІСТИЧНИХ ШКАЛ

Н.А. Яремчук[1], Р.С. Семенюк[2]

[1,2]Національний Технічний Університет України «Київський політехнічний інститут імені Ігоря Сікорського»
Київ, 03056, проспект Перемоги, 37, Україна
[1]yaremchukna@i.ua, [2]semeniukroman@i.ua

Анотація

У доповіді наведено основні аспекти проблеми побудови лінгвістичних шкал при застосуванні метричної класифікації в інтелектуальних вимірювальних системах. Результати вимірювання, які отримують на виході вимірювальних каналів системи, перетворюють в форму, що може бути застосована надалі при нечітких обчисленнях, прийнятті рішень, отриманні діагнозу про стан об'єктів, а також при необхідності регулювання параметрів об'єктів. Для отримання кінцевого результату використовується метрична класифікація, за якої на метричному носієві будується шкала з нечіткою лінгвістичною змінною, що дозволяє визначити клас еквівалентності якому відповідає результат вимірювання.

Показано, що при встановленні лінгвістичних шкал необхідно враховувати наступні складові невизначеності: невизначеність від нечіткості семантичного правила, за яким визначаються границі терм-множини лінгвістичної змінної, невизначеність від неповної ідентифікації об'єкта вимірювання (дефініціальна невизначеність); невизначеність від нестабільності латентного параметра, що підлягає вимірюванню; інструментальна невизначеність.

Встановлено, що при побудові лінгвістичних шкал може бути використано два способи врахування невизначеності, що відповідають в свою чергу двом алгоритмам роботи нечіткого класифікатора. Перший спосіб полягає в тому, що складові невизначеності, які стосуються шкали і об'єкту вимірювання, використовуються для побудови шкали, а інструментальна складова невизначеності використовується при оформленні результату вимірювання як нечіткого числа. Тоді за алгоритмом роботи нечіткого класифікатора знаходять композицію функцій приналежності терм-множини лінгвістичної змінної і нечіткого числа, а за центром ваги результату композиції знаходять необхідний клас еквівалентності. Другий спосіб полягає в знаходженні сумарної невизначеності за всіма складовими і її врахуванні при побудові лінгвістичної шкали. Тоді клас еквівалентності визначають за максимумом перерізу функцій приналежності терм-множини лінгвістичної шкали з результатом вимірювання, або за максимумом і мінімумом перерізу. Найчастіше для вирішення подібної задачі використовують нечіткий додаток до пакету MatLab, за яким визначають саме активовані класи еквівалентності, функції приналежності яких модифікують у відповідності з максимумом і мінімумом перерізу. Тому авторами доповіді був обраний другий спосіб урахування невизначеності і проаналізований зв'язок між індексом нечіткості функцій приналежності окремих термів лінгвістичної змінної і сумарною невизначеністю.

Ключові слова: невизначеність вимірювання, лінгвістична шкала, метрична класифікація

.

Постановка проблеми.

Сучасні інтелектуальні вимірювальні системи (ІВС) працюють разом з системами прийняття рішень, системами визначення стану об'єктів, системами регулювання параметрів об'єктів, тощо. Тому результати вимірювань, які отримують на виході вимірювальних каналів ІВС, перетворюють в форму, що може бути застосована надалі при нечітких обчисленнях. Для отримання кінцевого результату в ІВС використовується метрична класифікація [1], за якої на метричному носієві будується шкала з нечіткою лінгвістичною змінною (ЛЗ), що дозволяє визначити клас еквівалентності, якому відповідає результат вимірювання. Тому вважають, що особливістю ІВС є поліморфізм, який полягає в відображенні вимірюваної величини декількома шкалами, а саме метричною шкалою і шкалою класифікації.

Як відмічають вчені, що працюють в області створення інтелектуальних систем [2,3], головною проблемою є сумісне використання таких категорій як «неясність», «невизначеність», «нечіткість». За [2] неясність, невизначеність або нечіткість – це типові приклади категорій, оперування з якими неможливе з використанням тільки однієї теорії нечітких множин або нечіткої логіки для їх правильного відображення в інженерній діяльності. Для цього треба вміти добре представляти за допомогою абстрактних моделей, що відображають семантику об'єкта, знання та здатності людини для вирішення поставлених задач. Семантика повинна бути описана за допомогою добре визначеного синтаксису. За протилежністю від абстракції, семантика повинна бути реалістичною і прагматичною, щоб за її допомогою можна було б справитись з конкретною ситуацією в реальному світі, в якому для вирішення реальних проблем потрібні конкретні, індивідуальні і специфічні знання. Цю теорію і методи автори [2,3] відносять до інженерного мистецтва, що забезпечує плідний підхід до проблем, що вирішуються на основі знань, і при яких використовуються і знання про результати вимірювань, і експертні знання, і евристика. До класів неясності у вирішенні інженерних задач автор [2] відносить невизначеність або випадковість, що оцінюється за допомогою теорії ймовірностей, і нечіткість, як наслідок суб'єктивності, що оцінюється за теорією нечітких множин і теорії суб'єктивних ймовірностей. До класу невизначеності при встановленні шкал з нечіткою ЛЗ

при метричній класифікації автор [4] відносить інструментальну невизначеність, що супроводжує результати вимірювань окремих вимірювальних каналів ІВС.

Тому, щоб об'єднати перелічені вище чинники, які впливають на розмитість окремих градацій шкали з нечіткою ЛЗ, необхідна розробка підходів і рекомендацій з дотриманням умов обумовленості, практичності (прагматичності) і реалістичності (рис.1).

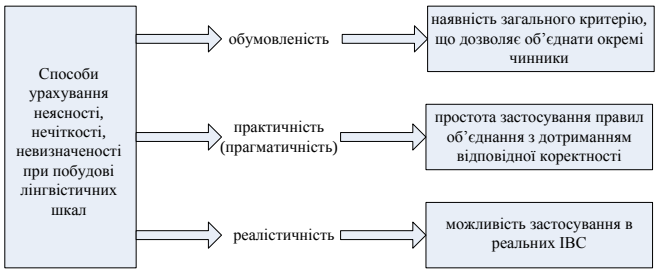

Рис. 1. Вимоги до способів урахування різних чинників при побудові лінгвістичних шкал в ІВС

Аналіз останніх досліджень і публікацій.

На основі досліджень, проведених авторами роботи [4] може бути запропонована наступна послідовність етапів встановлення або відтворення лінгвістичної шкали:

- орієнтовне визначення кількості термів множини ЛЗ і їх границь;
- аналіз нечіткості семантичного правила і складових невизначеності вимірювання;
- вибір загального критерію, що дозволяє об'єднати складові нечіткості і невизначеності;
- вибір алгоритму об'єднання складових нечіткості і невизначеності;
- перевірка максимальної кількості термів множини ЛЗ за умовою розрізненості;
- вибір функцій приналежності шкали з нечіткою ЛЗ;
- апробація (при наявності еталонних елементів) або моделювання нечіткого виведення з застосуванням встановленої лінгвістичної шкали.

При відтворенні лінгвістичної шкали спочатку визначають орієнтовну кількість термів (розрізнюваних градацій шкали), їх імена і границі у відповідності з існуючими рекомендаціями, настановами, нормативними документами, тощо. Якщо такої інформації немає, використовуються інтуітивні функції приналежності (ФП), які на підсвідомому рівні використовує людина [5]. Для того, щоб експертним шляхом визначити ширину терма, можна скористатись поняттям критичної точки ФП терма, під якою розуміють точку зі ступенем приналежності, що дорівнює 0,5 [5].

На другому етапі проводять аналіз чинників, які повинні враховуватись при побудові лінгвістичної шкали (рис.2).

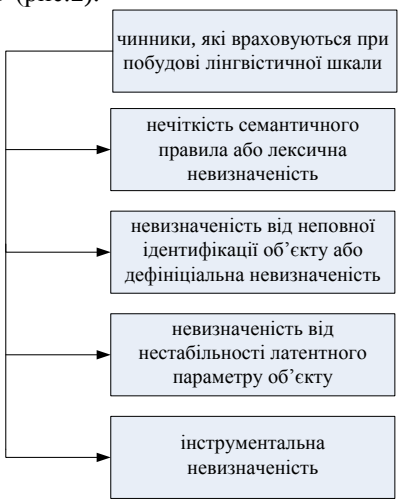

Рис. 2. Чинники, які враховують при побудові лінгвістичної шкали

Лінгвістичні шкали складаються з якісних оцінок фізичної величини, стану об'єктів або систем. Формалізація якісних оцінок ускладнена по-перше лінгвістичною нечіткістю понять (наприклад, «малий», «невеликий», а по-друге суб'єктивністю сприйняття цих понять різними експертами. За [2,3] семантичні правила – це правила, за якими об'єкти відносять до того чи іншого класу еквівалентності за набором їх ознак. В свою чергу семантичні правила складаються з правил інтерпретації і правил семантичного виведення для отримання загальних семантичних категорій. Правило інтерпретації включає ім'я класу еквівалентності, інтервал значень властивості або ознаки, що відповідає класу еквівалентності, а також математичну модель ступенів приналежності цьому класу, яка в більшості випадків є неточною або нечіткою. У зв'язку з цим в [5] введено поняття двох видів невизначеності: стохастичної і лексичної. Стохастична невизначеність означає невизначеність появи події, що є сама по собі точно описана. Лексична невизначеність означає невизначеність в описі події.

Для ідентифікації об'єкту, що підлягає дослідженню, необхідно мати опис його властивостей різної розмірності. При цьому ступінь розрізнення осіб або об'єктів може не завжди збільшуватись зі збільшенням розмірності опису властивостей. Тому на практиці залишають найбільш інформативні ознаки чи властивості, тим самим зменшуючи розмірність опису об'єкта. Наслідком цього є поява складової невизначеності від неповної ідентифікації об'єкта.

Якщо звернутись до [6], то невизначеність від неповної ідентифікації об'єкта відповідає дефініціальній невизначеності вимірювання, що виникає внаслідок обмеженої кількості деталей у визначенні вимірюваної величини (Definitional uncertainty – component of

measurement uncertainty resulting from the finite amount of details in the definition of measurement).

Якщо в процесі вимірювання параметри об'єкта можуть випадковим чином змінюватись, виникає складова невизначеності, пов'язана з непередбачуваними змінами вимірюваної величини. А похибки вимірювальних каналів ІВС призводять до появи інструментальної складової невизначеності.

Розглянуті вище чинники повинні бути враховані при побудові лінгвістичної шкали. Тому метою статті є розробка підходів до урахування нечіткості правил віднесення до певних класів еквівалентності і складових невизначеності вимірювання при побудові лінгвістичних шкал, які відповідають умовам наявності загального критерію, що дозволяє об'єднати окремі чинники, є простими в застосуванні і які забезпечують можливість застосування в реальних ІВС.

Виклад основного матеріалу дослідження.

Встановлено, що при побудові лінгвістичних шкал може бути використано два способи врахування невизначеності, що відповідають, в свою чергу, двом алгоритмам роботи нечіткого класифікатора. Перший спосіб полягає в тому, що складові невизначеності, які стосуються шкали і об'єкту вимірювання, використовуються для побудови шкали, а інструментальна складова невизначеності використовується при оформленні результату вимірювання як нечіткого числа (Fuzzy number [7]). Тоді значення функції приналежності результату вимірювання x_p дорівнює одиниці, тобто $\mu(x_p)=1$, а інструментальна невизначеність використовується для оформлення розмитості нечіткого числа. Тоді за алгоритмом роботи нечіткого класифікатора знаходять композицію функцій приналежності терм-множини лінгвістичної змінної і нечіткого числа, а за центром ваги результату композиції знаходять клас еквівалентності шкали, якому відповідає x_p.

Другий спосіб полягає в знаходженні сумарної невизначеності за всіма чинниками і її врахування при побудові лінгвістичної шкали. Тоді результат вимірювання оформлюють як singleton [7], тобто

$$\mu(x)=\begin{cases}1, \textit{ якщо } x=x_p;\\ 0 \textit{ в інших випадках.}\end{cases}$$

Клас еквівалентності визначають за максимумом перерізу функцій приналежності терм-множини лінгвістичної шкали з результатом вимірювання або за максимумом і мінімумом перерізу. Найчастіше для вирішення подібної задачі використовують нечіткий додаток до пакету MatLab, за яким визначають саме активовані функції приналежності класів еквівалентності. Ці ФП модифікують у відповідності з максимумом і мінімумом перерізу і за центром ваги визначають клас еквівалентності. Тому авторами статті був обраний другий спосіб урахування чинників, що характеризують нечіткість і невизначеність.

Як загальний критерій, що дозволяє об'єднати наведені на рис.2 чинники, було обрано індекс нечіткості, що є мірою нечіткості множини, визначеною як відстань цієї множини до найближчої до неї звичайної множини і яка задовольняє всім аксіомам метрик незалежно від того яка з них (лінійна або евклідова) при цьому використана [8]. Ця метрика може бути застосована і для дискретної і для неперервної множини, тобто при всіх способах подання нечіткості семантичного правила.

Лінійний індекс нечіткості неперервної множини $\underset{\sim}{A}$ з $\mu_{\underset{\sim}{A}}(x)$ визначається як

$$I_A^L=\frac{2}{b-a}\int_a^b|\mu_{\underset{\sim}{A}}(x)-\mu_{\underline{\underset{\sim}{A}}}(x)|\,dx, \qquad (1)$$

де $\mu_{\underline{\underset{\sim}{A}}}(x)$ – звичайна множина, найближча до нечіткої, а a і b – нижня і верхня границі носія нечіткої множини $\underset{\sim}{A}$ відповідно.

Дослідження зв'язку індексу нечіткості з частиною носія, що характеризує нечіткість семантичного правила було проведено для різних форм ФП. Для прикладу обрано клас еквівалентності «малий», тобто початковий терм лінгвістичної шкали з ФП

$$\mu_{T_1}(x)=\begin{cases}1, \textit{ якщо } 0\le x\le a_1;\\ \dfrac{a_2-x}{a_2-a_1}, \textit{ якщо } a_1\le x\le a_2;\\ 0, \textit{ якщо } a_2<x,\end{cases}$$

де $a=0;\ b=a_2;\ 0\le x\le a_1$ – область толерантності, де $\mu_{T_1}(x)=0;\ a_1\le x\le a_2$ - область, яка характеризує нечіткість семантичного правила; значення $(a_1+a_2)/2$ відповідає чіткій верхній границі класу еквівалентності «малий»; $(a_2-a_1)/2$ - область розмитості границі.

Якщо ввести позначення $\Delta=(a_2-a_1)/2$ і $\delta=2\Delta/(a_1+a_2)$, δ - відносна розмитість границі, тоді $I_A^L=\delta/(1+\delta)$.

Для індексу нечіткості, заснованого на евклідовій відстані

$$I_A^E=\frac{2}{b-a}\sqrt{\int_a^b(\mu_A(x)-\mu_{\underline{\underset{\sim}{A}}}(x))^2dx}=\sqrt{\frac{2\delta}{(1+\delta)3}} \qquad (2)$$

Виходячи з (1) і (2) при об'єднанні чинників, що впливають на індекс нечіткості, треба визначити загальне значення області розмитості Δ, як носія нечіткого числа, симетричного відносно границі терму або класу еквівалентності $(a_1+a_2)/2$. Подібна задача вирішується з застосуванням теорії можливостей і теорії нечітких множин при об'єднанні результатів вимі-

рювання, отриманих за допомогою «фізичних» систем і результатів оцінювання з «суб'єктивною» неточністю [9,10]. В цих роботах розглянуто інтервальний аналіз, за яким пропонується дуже простий розрахунок покривного інтервалу з ймовірністю 100%. Але ця оцінка є завищеною і тому в більшості випадків використовується покривна ймовірність 0,95 або 0,99, що відповідає множині α - зрізу ФП $\alpha = 0,5$ або $\alpha = 0,01$.

Як приклад аналізу складових нечіткості і невизначеності можна навести процедуру встановлення лінгвістичної шкали артеріального тиску в ІВС визначення допустимого рівня фізичного навантаження залежності від стану здоров'я людини [11]. Допустимий рівень фізичного навантаження задано лінгвістичною шкалою: «extra light» «very light», «light», «moderate», «hard», «maximum», і для визначення необхідного класу еквівалентності за цією шкалою використовується система нечіткого виведення з використанням лінгвістичних шкал артеріального тиску (систолічного і діастолічного) і пульсу. При встановленні шкали артеріального тиску терм-множина лінгвістичної змінної визначається за експертними даними. Розбіжність даних, отриманих з різних джерел для певного віку і статі особи, ідентифікується як нечіткість семантичного правила. Наприклад, для терму «нормальний» вона складала ± 3 *мм рт. ст.*. Для повної ідентифікації об'єкта треба було розробити шкали з урахуванням статі і віку (по 3 роки в кожній групі). Це збільшувало об'єм бази правил і, відповідно, кількість антецедентів в ній. Тому треба було знайти компроміс між критеріями точності при побудові шкали і критеріями компактності бази правил [12]. Виходячи з подальших чинників, що впливають на індекс нечіткості, а саме інструментальної невизначеності ± 3 *мм рт. ст.* і нестабільності латентного параметру ± 5 *мм рт. ст.* (від зміни тиску протягом доби і впливу побутових стимуляторів (кави, чаю)), залишкову складову невизначеності від неповної ідентифікації об'єкту було встановлено на рівні $\pm 2,5$ *мм рт. ст.* з організацією шкал для трьох вікових груп: «молодий», «середній», «похилий». У відповідності з рекомендаціями [9,10] було визначено загальний покривний інтервал ± 10 *мм рт. ст.* за покривною ймовірністю $P = 0,95$.

Висновки

Наведено основні аспекти побудови лінгвістичних шкал в інтелектуальних вимірювальних системах при застосуванні метричної класифікації, за якої на метричному носієві будується шкала з нечіткою лінгвістичною змінною, що дозволяє визначити клас еквівалентності якому відповідає результат вимірювання.

На основі аналізу літературних джерел показано, що однією з головних проблем встановлення лінгвістичної шкали є сумісне використання таких категорій як неясність, невизначеність, нечіткість. Вони обумовлюють наявність чинників, які впливають на розмитість окремих градацій лінгвістичної шкали. У зв'язку з цим виникає необхідність в розробці способів урахування неясності, нечіткості, невизначеності при побудові лінгвістичних шкал в ІВС, які відповідають вимогам обумовленості, прагматичності і реалістичності.

В роботі запропонована послідовність етапів встановлення або відтворення лінгвістичної шкали, причому основна увага приділена етапам аналізу основних чинників, що впливають на розмитість шкали і вибору загального критерію, на основі якого можна об'єднати ці чинники.

Встановлено, що чинники, які враховуються при побудові лінгвістичної шкали – це нечіткість семантичного правила, за яким встановлюються терми ЛЗ, невизначеність від неповної ідентифікації об'єкта або дефініціальна невизначеність, невизначеність від нестабільності латентного параметру об'єкту вимірювання, інструментальна невизначеність.

Встановлено, що при побудові лінгвістичних шкал може бути використано два способи врахування невизначеності, що відповідають, в свою чергу, двом алгоритмам роботи нечіткого класифікатора. З них обрано спосіб, орієнтований на використання нечіткого MatLab, за яким при побудові лінгвістичної шкали об'єднують усі чинники.

На основі результатів дослідження як загальний критерій було обрано індекс нечіткості терм-множини лінгвістичної змінної з використанням носія множини у вигляді інтервалу, що характеризує область нечіткості. Покривний інтервал може бути розрахований з використанням співвідношень, що розроблені вченими, які працюють в області застосування теорії можливостей і нечітких множин.

Наведено приклад реальної ІВС з аналізом чинників, які враховуються при побудові лінгвістичної шкали.

Список літератури

1. Кендалл М., Стьюарт А. Статистические выводы и связи. Пер. с англ. Изд. «Наука», М.: 1973. 899с.
2. Танака К. итоги рассмотрения факторов неопределенности и неясности в инженерном искусстве. В книге «Нечеткие множества и теория возможностей» под ред. Р. Ягера. Пер. с англ. Москва «Радио и связь». 1986. стр 37-50.
3. Руспини Э.Г. Последние достижения в нечетком кластер-анализе. В книге «Нечеткие множества и теория возможностей» под ред. Р. Ягера. Пер. с англ. Москва «Радио и связь». 1986. Стр 114-132.
4. Семенюк Р.С., Яремчук Н.А. Основні етапи встановлення лінгвістичних шкал при вимірюваннях і діагностиці. Матеріали VII Міжнародної науково-технічної конференції «Метрологія, інформаційно-вимірювальні технології та системи (МІВТС)», лютий 18, 19 2020, Харків.

5. Пегат А. Нечеткое моделирование и управление. Пер. с англ. М.: БИНОМ. Лаборатория знаний, 2009. 798 с.
6. JCGM 200: 2008. International vocabulary of metrology – Basic and general concepts and associated terms (VIM). JCGM 2008
7. Зак Ю.А. Принятие решений в условиях нечетких и размытых данных: Fuzzy – технологии. М.: Книжный дом «ЛИБРОКОМ», 2013. 352с.
8. Конішева Л.К., Назаров Д.М. Основы теории нечетких множеств: Учебное пособие. – СПб.: Питер. 2011. 192с.
9. Mauris G. Representing and Approximating Symmetric and Asymmetric Probability Coverage Intervals by Possibility Distributions. IEEE Transactions on Instrumentation and Measurement. Vol. 58, N1 2009. pp. 41-45.
10. Ferrero A., Prioli M., Salicone S. Conditional Random-Fuzzy Variables Representing Measurement Results. IEEE Transactions on Instrumentation and Measurement. Vol. 64. N5, 2015 pp. 1170-1178.
11. Tóth-Laufer E., Várkonyi-Kóczy A. A Soft Computing-Based Hierarchical Sport Activity Risk Level Calculation Model for Supporting Home Exercises. IEEE Transactions on Instrumentation and Measurement. Vol. 63, N6, 2014 pp. 1400.
12. Штовба С.Д., Штовба О.В., Панкевич О.Д. критерії точності та компактності для оцінювання якості нечітких баз знань у задачах ідентифікації. Наукові праці ВНТУ, 2012, N4.

References

1. Kendall, Stuart 1973 – The Advanced Theory of Statistics, 723 pp.
2. Kazuo Tanaka, Hua O. Wang Fuzzy Control Systems Design and Analysis: A Linear Matrix Inequality Approach. 2001. John Wiley &Sons, INC. ISBNS: 0-471-32324-1 (Hardback).-305p.
3. Enrique H. Ruspini, A new approach to clustering, Information and Control, Volume 15, Issue 1, 1969, Pages 22-32, ISSN 0019-9958, https://doi.org/10.1016/S0019-9958(69)90591-9.
4. Семенюк Р.С., Яремчук Н.А. Основні етапи встановлення лінгвістичних шкал при вимірюваннях і діагностиці. Матеріали VII Міжнародної науково-технічної конференції «Метрологія, інформаційно-вимірювальні технології та системи (МІВТС)», лютий 18, 19 2020, Харків.
5. Piegat, Andrzej. (2001). Fuzzy Modeling and Control. 10.1007/978-3-7908-1824-6. (Пегат А. Нечеткое моделирование и управление. Пер. с англ. М.: БИНОМ. Лаборатория знаний, 2009. 798 с).
6. JCGM 200: 2008. International vocabulary of metrology – Basic and general concepts and associated terms (VIM). JCGM 2008
7. Zak Yu. A. (2013), Decision-making under fuzzy and blurry data: Fuzzy-technology, Librokom, Moscow, 352 p. (rus).
8. Konysheva L.K. Osnovy teorii nechetkih mnozhestv: Uchebnoe posobie [Fundamentals of the theory of fuzzy sets: a Training manual] / L.K.Konysheva, D.M.Nazarov. - SPb.: Piter, 2011. - 192 p. [in Russian]
9. Mauris G. Representing and Approximating Symmetric and Asymmetric Probability Coverage Intervals by Possibility Distributions. IEEE Transactions on Instrumentation and Measurement. Vol. 58, N1 2009. pp. 41-45.
10. Ferrero A., Prioli M., Salicone S. Conditional Random-Fuzzy Variables Representing Measurement Results. IEEE Transactions on Instrumentation and Measurement. Vol. 64. N5, 2015 pp. 1170-1178.
11. Tóth-Laufer E., Várkonyi-Kóczy A. A Soft Computing-Based Hierarchical Sport Activity Risk Level Calculation Model for Supporting Home Exercises. IEEE Transactions on Instrumentation and Measurement. Vol. 63, N6, 2014 pp. 1400.
12. Shtovba S.D., Shtovba O.V., Pankevich O.D. Accuracy and compactness criteria for evaluating the quality of fuzzy knowledge bases in identification problems // Scientific Works of Vinnytsia National Technical University. – 2012. – №4.