

МОДЕЛЬ ДВІЙКОВОГО МЕТОДУ ДЕРЕВОПОДІБНОЇ КЛАСТЕРИЗАЦІЇ ВЛАСТИВОСТЕЙ ЯКОСТІ ІНФОРМАЦІЙНОЇ СИСТЕМИ

Vavilov Ye. V. Model of binary method tree-like clustering properties of the quality of the information system. The article is based on a generalized solution of automatic generation of quality measures and information system model of construction wood properties of an information system given solution development method binary encoding genetic information system properties. The method is based on principles that are characteristic of genetic algorithms. Technology of wood properties for the information system is presented as a classification algorithm, selection and grouping a set of quality indicators. These indicators can be used to build levels of nodes and leaves of the tree.

Keywords: informative system, quality indicators, wood properties, genetic information, genetic algorithm, descriptions model

Вавілов Є. В. Модель двійкового методу деревоподібної кластеризації властивостей якості інформаційної системи. В статті приведено рішення задачі розробки методу двійкового кодування генетичної інформації властивостей інформаційної системи. Метод засновано на принципах, які є характерними для генетичних алгоритмів. Технологія побудови дерева властивості для інформаційної системи представлено у вигляді алгоритму класифікації, відбору та групування набору показників якості.

Ключові слова: інформаційна система, показники якості, дерево властивостей, спадкова інформація, генетичний алгоритм, модель характеристик

Вавілов Е. В. Модель двоичного метода древовидной кластеризации свойств качества информационной системы. В статье приведено решение задачи разработки метода двоичного кодирования генетической информации свойств информационной системы. Метод основан на принципах, которые являются характерными для генетических алгоритмов. Технологию построения дерева свойства для информационной системы представлено в виде алгоритма классификации, отбора и группировки набора показателей качества.

Ключевые слова: информационная система, показатели качества, дерево свойств, наследственная информация, генетический алгоритм, модель характеристик

1. Постановка задачі

Ускладнення об'єктів вимірювання в сучасних інформаційних системах (СІМ) та мережах призвело до створення інформаційно-вимірювальних систем (ІВС), які, відповідно до ГОСТ 8.437-81, функціонують у складі комплексів контролю за якістю, як сукупність функціонально об'єднаних вимірювальних, обчислювальних та інших допоміжних технічних засобів для отримання вимірювальної інформації, її перетворення, обробки з метою подання споживачеві в необхідному вигляді, або автоматичного здійснення логічних функцій контролю, діагностики та ідентифікації в СІМ. В даний час велика увага приділяється ІВС, які здатні функціонувати в умовах мінливих характеристик каналів зв'язку, а також адаптуватися до різних топологій та методів передачі вимірювальної інформації, представленої в цифровому вигляді, що, зокрема, має важливе значення при побудові територіально-розподілених вимірювальних систем, та є однією із заporук успіху наукового, технічного та технологічного прогресу щодо отримання найповніших даних про досліджувані об'єкти.

Все це відзначається багатьма вченими та узагальнено, наприклад, у [1]. Зважаючи на це та на публікацію автора [2], враховуючи високу складність ІВС та методів, що забезпечують достовірне представлення даних про об'єкти вимірювань, вимог щодо їх метрологічного забезпечення, *метою статті* є побудова моделі двійкового методу деревоподібної кластеризації властивостей якості інформаційної системи на основі двійкового кодування генетичної інформації властивостей інформаційної системи. В основу розробки покладено

той факт, що метод засновано на принципах, які є характерними для генетичних алгоритмів (ГА). Це надасть можливості удосконалити критерії оптимальності якісної роботи інформаційної системи, які засновано на принципах кваліметрії.

Зв'язок задачі з важливими науковими та практичними завданнями. Вимоги до окремих аспектів забезпечення якості, управління та структура показників якості (ПЯ) регламентуються серією стандартів ISO-9000. Проблема комплексної оцінки й оптимізації ПЯ СІМ повинна вирішуватися в рамках цих аспектів. Важливу роль при цьому відіграє й оптимальне поєднання ціни виробу, його проектування, виробництва, і якості. Необхідність комплексної оцінки якості СІМ на етапах життєвого циклу висуває ряд питань.

1) Структура узагальнених ПЯ, яка, природно, повинна будуватися на встановлених стандартами ПЯ технічних засобів та видів забезпечення (програмного, інформаційного) та їх зв'язок з процесами проектування та виробництва.

2) Інтеграція критеріїв оптимальності якості СІМ.

3) Зв'язок комплексних (інтегральних) ПЯ з характеристиками процесів проектування, виробництва та експлуатації.

Сукупність поставлених питань веде до необхідності введення багатокритеріальних та багатопараметричних оптимізаційних задач. Як покано у багатьох першоджерелах, в останні роки для вирішення подібних завдань з метою пошуку оптимальних характеристик та формулювання критеріїв оптимальності СІМ на протязі життєвого циклу, інтенсивно використовується апарат генетичних алгоритмів.

Як видно, мета статті відповідає зазначеним положенням і, т.ч., **раніше невирішеною частиною загальної проблеми** є задача автоматичного формування системи ПЯ, яка б базувалася на моделі двійкового методу деревоподібної кластеризації властивостей якості інформаційної системи. Тоді, згідно до неї, було б можливим провести оцінку функціонування СІМ у відповідності до встановлених вимог.

Аналіз останніх досліджень і публікацій, в яких покладено початок вирішення проблеми. Як зазначено автором у [2], О. Грабовський у [1] та інші вчені, наприклад, у [3, 4], показали, що сукупність питань з удосконалення метрологічного забезпечення у галузі ІС, веде до необхідності введення багатокритеріальних та багатопараметричних оптимізаційних задач [5, 6]. Так, О. Лужбінін вважає, що в останні роки для вирішення подібних завдань з метою пошуку оптимальних характеристик та формулювання критеріїв оптимальності ІВС, став інтенсивно використовуватися апарат, заснований на теорії ГА. Він зазначає, що у цьому напрямі перші публікації належать Н. Барічеллі, Л. Фогелю, А. Оуенсу та М. Уолшу. В той же час загальновідомо, що сучасну теорію ГА розробив Д. Холланд, яку в подальшому удосконалили його учні – К. де Йонг та Д. Голдберг.

Розвитком теорії та їх практичним застосуванням у сфері програмно-технічних засобів займалися О. Лужбінін та С. Ісаєв. Г. Вороновський, К. Махота, С. Петраш та С. Сергєєв, які показали можливість використання ГА при дослідженні штучних нейронних мереж. В. Емельянов та В. Курейчик узагальнили відомі положення щодо теорії та практики еволюційного моделювання, та разом з О. Лебедевим заклали основи пошукової адаптації.

Д. Рутковська, М. Пилиньський та Л. Рутковський показали, що ГА можуть бути основним засобом дослідження нечітких систем, що лягло в основу досліджень у дисертаційній роботі. В області оцінки якості електронної апаратури необхідно відзначити російських вчених В. Брюніна, В. Абрамова, В. Пролейко та Б. Мансурова.

В області оцінки якості програмних засобів відомими є роботи Є. Кранкова, В. Ліпаєва та М. Треногіна. В області оптимізації систем на основі ГА основними є роботи Д. Батищева та І. Норенкова. Однак, слід зазначити, що у зазначених напрямках досліджень для вирішення завдань удосконалення моделей двійкових методів деревоподібної кластеризації

властивостей якості інформаційної системи на основі двійкового кодування генетичної інформації властивостей інформаційної системи, питання у належній мірі не розроблено, хоча відомі дослідження за спільним рішенням деяких окремих завдань.

2. Поняття та визначення

У [2] йде мова про те, що ПЯ будь якої технічної системи, зокрема, такої, як СІМ, можна представити у вигляді багатомірної структурованої множини, лінеаризація якої призводить до формування та побудови дерева властивостей [7-9]. Вважатимемо, що термін «дерево властивостей» є тотожним терміну «дерево характеристик».

Зважаючи на це, результати, які приведено у [3-21], можна використати для множинного визначення дерева властивостей, що буде еквівалентним побудові дерева характеристик якості СІМ підприємства [8-22]. У цих же джерелах приведено основні положення, методи та алгоритми процесу створення багатомірної множини з її послідовним перетворенням у дерево властивостей. Враховуючи їх, побудову дерева властивостей якості СІМ можна представити у вигляді алгоритму класифікації, відбору та групування множини ПЯ, що зроблено автором у [2]. Зважаючи на результат, отриманий у [2], розробимо модель двійкового методу деревоподібної кластеризації.

3. Модель двійкового методу деревоподібної кластеризації

З метою розробки моделі двійкового методу деревоподібної кластеризації врахуємо, що запропонований у [7] метод невиваженого попарного арифметичного середнього (англ.: *Unweighted pair-group method using arithmetic averages*, UPGMA) полягає в обчисленні відстані між двома різними кластерами, як середнього арифметичного значення між усіма парами об'єктів у них. Цей метод доволі широко використовується у філогенетичному аналізі. Суттю його є аналіз та кластеризація за загальними ознаками, а також визначення відстаней та побудова гілок та вузлів дерева. Метод себе добре зарекомендував при дослідженнях змін генотипу, які відбуваються під впливом випадкових схрещувань та генних мутацій.

Однак, якщо формування генотипу особин виконувати за наперед зазначеними правилами, тобто виключити випадковість, метод UPGMA стає надлишковим. Це приводить до висновку про те, що його можна замінити більш простими методами. Один з таких методів розроблюється та пропонується нижче.

Приведений вище метод двійкового кодування спадкової інформації властивостей системи має на увазі виключення випадковості, оскільки ідентифікаційний код особи визначається виразом, який згідно до [2], є ідентифікатором особи:

$$\{chq_x^N\} = \left\{ \left\{ chq_{x_{N-1}}^{N-1} \right\}, \left\{ \beta_N^{Q_x^N} \right\} \right\}; \quad (N \geq 1),$$

де: $\{chq_{x_{N-1}}^{N-1}\}$ – хромосома ідентифікатора предка в шарі $(N-1)$;

x_{N-1} – ідентифікатор предка в шарі $(N-1)$;

$\{\beta_N^{Q_x^N}\}$ – ген ідентифікатора особи Q_x^N .

Зазначимо, що пояснення щодо використаних понять та позначень є достатньо об'ємними, та приведені автором у [2]. Там же показано, що нормування довжини $l_{x_N}^N$ гена $\beta_N^{Q_x^N}$ в межах шару можна здійснювати на основі виразу:

$$l_{norm}^N = \max |Q^{N-1}|, \quad (1)$$

де $\max |Q^{N-1}|$ – потужність одного з множин шару $N-1$, що має максимальне значення.

Т.ч., хромосома-нащадок у якості ознаки приналежності до кластеру предка містить код його хромосоми. При такому формуванні процес кластеризації можна уявити як послідовний

перебір хромосом та порівняння кодів складових їх генів з ознакою кластера. Оскільки генна інформація кодується двійковим кодом, то таке порівняння досить просто здійснити з використанням логічної операції кон'юнкції. Рівність результату операції нулю визначає приналежність хромосоми до кластера:

$$ch_x^N \in C_i \left\{ \left\{ ch_i \{ sequence_N \} \right\} \wedge \left\{ ch_x^N \right\} = 0, \right. \quad (2)$$

де C_i – i -кластер;

ch_i – признак i -кластеру;

$sequence_N$ – послідовність (множина) бітів довжиною

$$l_{sequence_N} = \left| ch_x^N \right| - \left| ch_i \right|, \quad (3)$$

всі елементи q_s якої дорівнюють 0 ($\forall q_s \in sequence_N, q_s := 0$).

Як приклад та підтвердження розроблених теоретичних положень, розглянемо модель формування кластерів для структури, коди хромосом якої наведено у Табл. 1.

Табл. 1

Коди хромосом	$l_{x_N}^N$	l_{norm}^N	$l_{sequence_N}$	$\{ch_i \{sequence_N\}\}$	Приналежність до кластера
101	2	2	0	101	Q1
110	2	2	0	110	Q2
101001	3	3	3	101000	Q1
110001	2	3	3	110000	Q2
101100	3	3	3	101000	Q1
101010	3	3	3	101000	Q1
110010	2	3	3	110000	Q2
101001001	2	3	6	101000000	Q1
110010010	2	3	6	110000000	Q2
110001001	3	3	6	110000000	Q2
110001010	3	3	6	110000000	Q2
101010100	3	3	6	101000000	Q1
101100001	2	3	6	101000000	Q1
101100010	2	3	6	101000000	Q1
101010001	3	3	6	101000000	Q1
101010010	3	3	6	101000000	Q1
110001100	3	3	6	110000000	Q2
110010001	2	3	6	110000000	Q2
101001010	2	3	6	101000000	Q1

У Табл. 1 встановлено, що ознакою кластера Q1 є $\{ch_1\} = \{101\}$, а кластера Q2 – $\{ch_2\} = 110$. Візуалізація формування кластерів приведено на Рис. 1, а...г. Зазначимо, що комп'ютерне моделювання виконувалося на підставі виразів (1)...(3) з використанням програмних середовищ LabVIEW та MatLab.

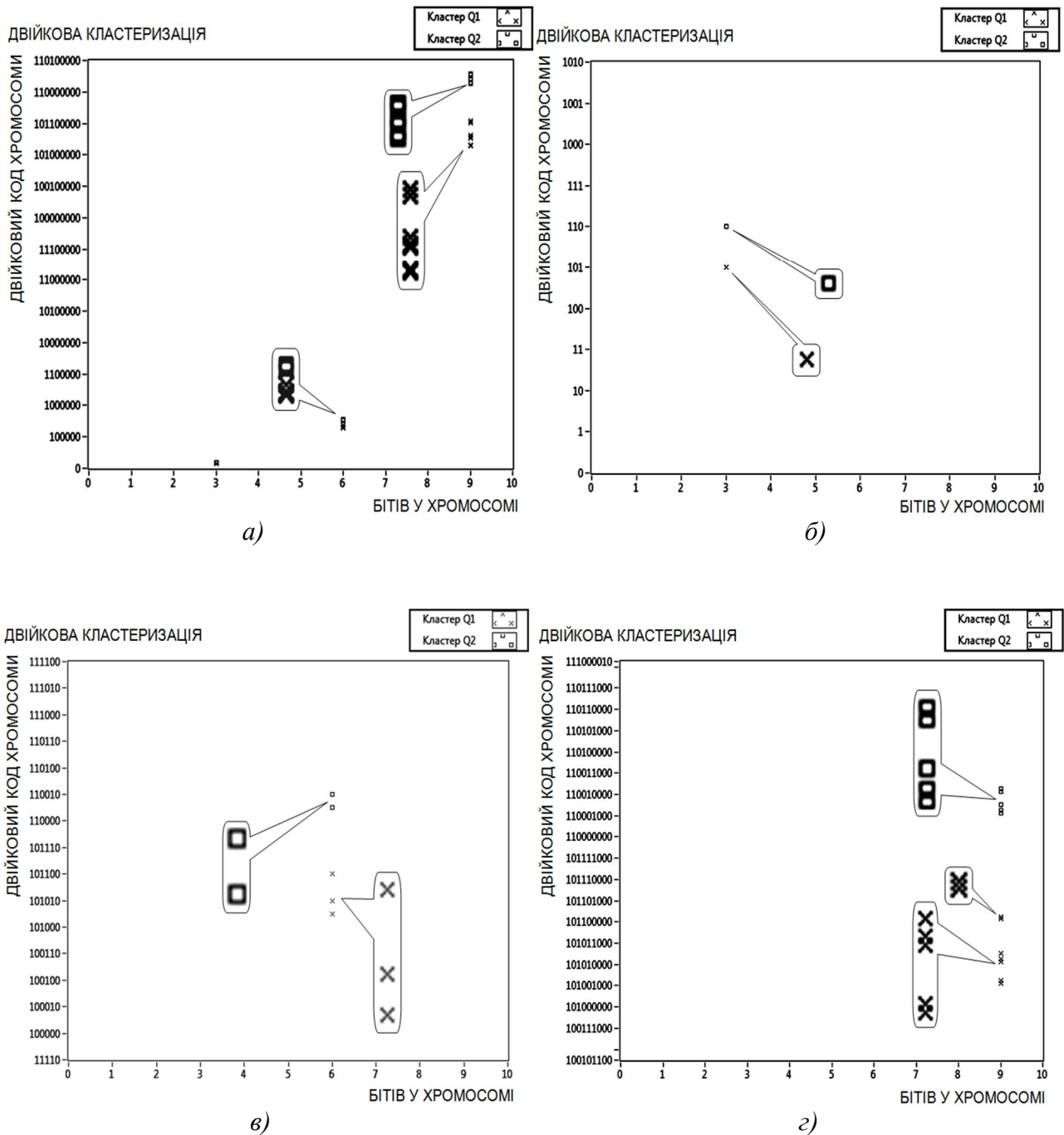


Рис. 1. Формування кластерів:

a – загальний вигляд кластера; *б, в, г* – деталізація кластерів по довжині хромосоми

Як видно з Рис. 1, запропонований спосіб генного кодування та кластеризації дає можливість автоматичного розподілу елементів кластерів за шарами, що визначаються нормованою довжиною хромосом (нормоване число бітів в генах, що складають хромосому).

На Рис. 2 наведено повне дерево структури з теоретичним коренем дерева $\{Q_1^0\}$, ознака кластера $Q_0 \{ch_0\} = \{1\}$. Запропонований метод дозволяє в такому ж автоматичному режимі формувати піддерева властивостей. На Рис. 3 та Рис. 4 наведено зображення піддерев, корінням яких є множини $\{Q_1^1\}$ та $\{Q_2^1\}$ з ідентифікаторами $\{101\}$ та $\{110\}$ відповідно.

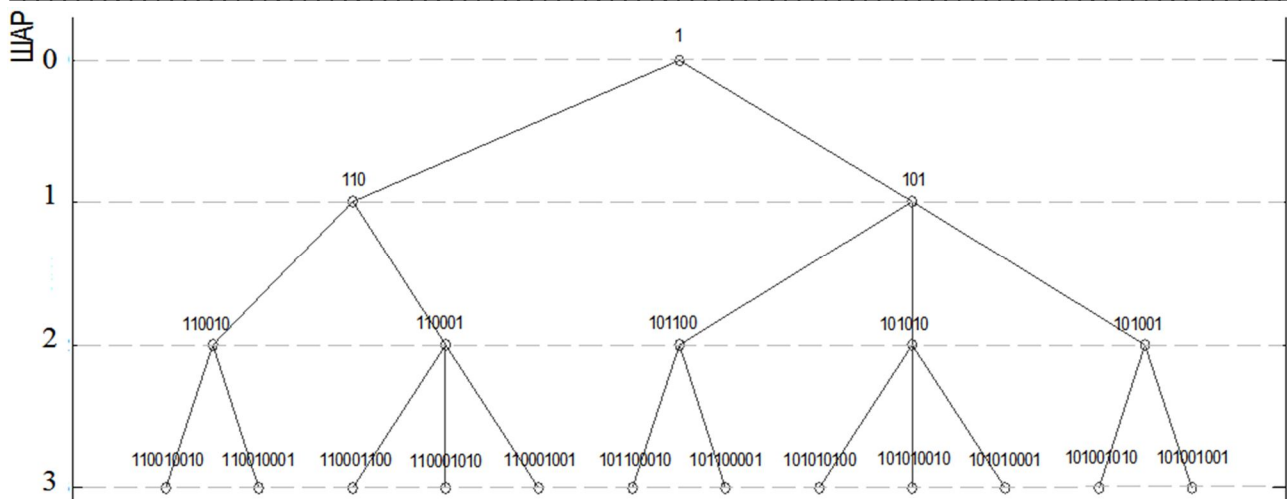


Рис. 2. Повне дерево структури з теоретичним коренем дерева $\{Q_0^0\}$ (кількість шарів – 3); ознакою кластера $Q_0 \in \{ch_0\} = \{1\}$

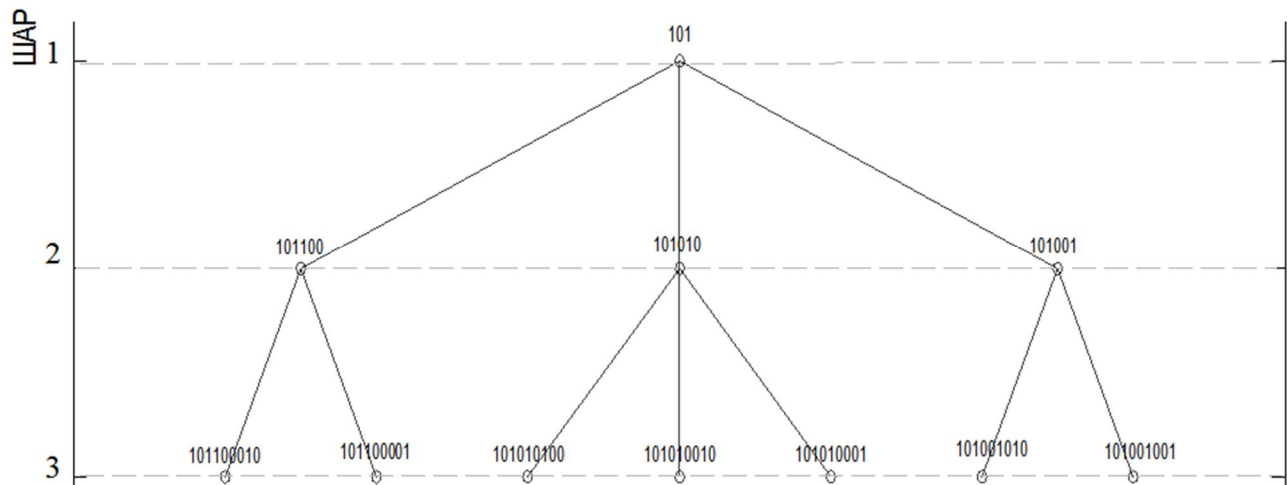


Рис. 3. Піддерево, коренем якого є множина $\{Q_1^1\}$ з ідентифікатором $\{101\}$ (кількість шарів – 2)

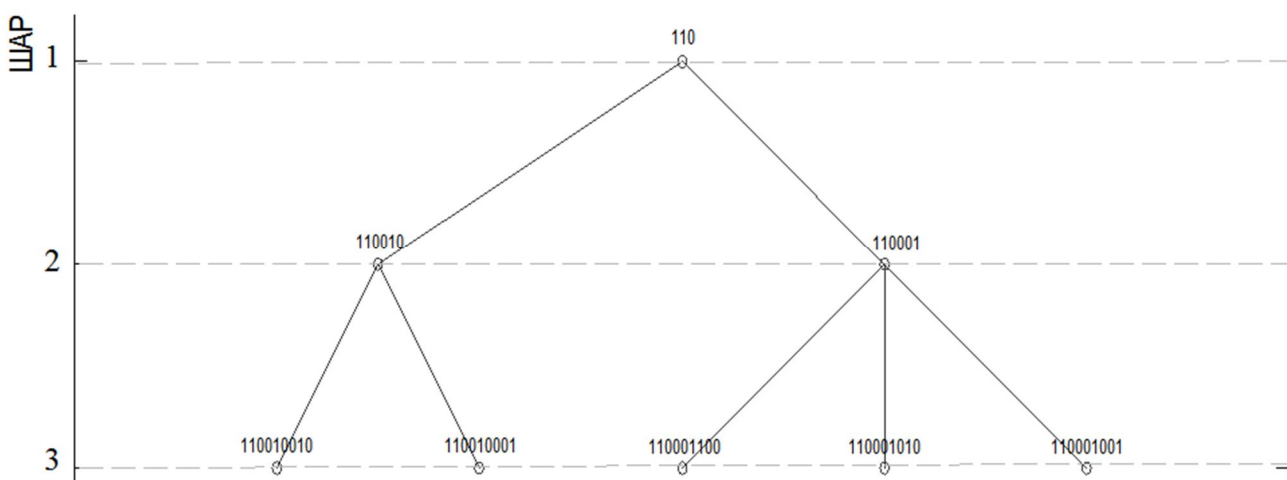


Рис. 4. Піддерево, коренем якого є множина $\{Q_2^1\}$ з ідентифікатором $\{110\}$ (кількість шарів – 2)

5. Висновки

Приведене рішення проблеми розробки методу двійкового кодування генетичної інформації властивостей системи, що засновано на принципах, які є характерними для генетичних алгоритмів, дозволяє побудувати дерево властивостей якості для інформаційної системи, що представлено у вигляді алгоритму класифікації, відбору та групування набору показників якості. Ці показники можуть бути використані для побудови рівнів вузлів та листків дерева.

Література

1. Грабовський О. В. Розробка методу оцінювання якості інформаційно-вимірювальних систем на основі використання генетичних алгоритмів : автореферат дис... канд. техн. наук : 05.01.02 / О. В. Грабовський; ОДАТРЯ. — Одеса, 2013. — 20 с.
2. Вавілов Є. В. Множинне визначення дерева властивостей якості інформаційної системи / Є. В. Вавілов // Наукові записки Українського науково-дослідного інституту зв'язку. — 2015. — №5(39). — С. 78-86.
3. Вавілов Є. В. Загальні принципи технології прогнозування ситуацій для систем з генетичними принципами функціонування / Є. В. Вавілов // Інформаційно-вимірювальні технології в метрології, технічне регулювання та менеджмент якості : III Всеукраїнська науково-практична конференція, 30-31 травня 2013 р. : матеріали конференції. — Одеса, ОДАТРЯ. — С. 213-215.
4. Вавілов Є. В. Технологія прогнозування ситуацій для систем з генетичними принципами функціонування / Є. В. Вавілов // Технологічний аудит та резерви виробництва. — 2013. — №2/1(10). — С. 22-24.
5. Грабовський О. В. Принципові питання вирішення задачі багатокритеріальної оптимізації показників якості інформаційно-вимірювальних систем на основі мультихромосомного генетичного алгоритму / О. В. Грабовський, Л. В. Коломієць // Збірник наукових праць Одеської державної академії технічного регулювання та якості. — 2013. — №1 (2). — С. 93-101.
6. Грабовський О. В. Технологія розв'язку завдання багатокритеріальної оптимізації показників якості ІВС на основі мультихромосомного генетичного алгоритму / О. В. Грабовський, Є. В. Вавілов // Сучасний стан та перспективи розвитку системи технічного регулювання, метрології та якості : VI Всеукраїнська науково-практична конференція молодих учених і студентів, 21-22 травня 2015 р. : матеріали конференції. — Одеса, ОДАТРЯ. — С. 11.
7. Грабовський О. В. Філогенетична модель побудови дерева властивостей показників якості / О. В. Грабовський // Вісник інженерної академії України. — 2013. — №1. — С. 225-228.
8. Грабовський О. В. Візуалізація структури показників якості інформаційно-вимірювальних систем / О. В. Грабовський, С. Л. Волков, О. О. Скопа // Метрологія та прилади. — 2013. — №2. — С. 69-74.
9. Грабовський О. В. Візуалізація структури показників якості інформаційно-вимірювальних систем / О. В. Грабовський, С. Л. Волков, О. О. Скопа // Інформаційно-вимірювальні технології в метрології, технічне регулювання та менеджмент якості : III Всеукраїнська науково-практична конференція, 30-31 травня 2013 р. : тези доповіді. — Одеса, 2013. — С. 74-75.
10. Грабовський О. В. Аналіз показників якості інформаційно-вимірювальних систем / О. В. Грабовський // Вісник національного технічного університету «ХПІ». — 2013. — №16. — С. 59-66.

11. Вавилов Е. В. Задача выбора оптимальной структуры системы метрологического обеспечения и управления / Е. В. Вавилов, О. Г. Есина // Проблемы технічного регулювання та якості : IV Всеукраїнська науково-практична конференція, 9-10 жовтня 2014 р. : тези доповідей. — Одеса, ОДАТРЯ. — С. 147-148.
12. Грабовський О. В. Скорочення випробувань надійності ІВС за рахунок її функціональної надмірності / О. В. Грабовський, Н. Ф. Казакова // Технологічний аудит та резерви виробництва. — 2013. — №2/1(10). — С. 24-27.
13. Вавілов Є. В. Інтелектуальні автономні системи: Концептуальні положення створення та функціонування / Є. В. Вавілов, О. О. Скопа // Біоніка інтелекту. — 2013. — №5(74). — С. 95-102.
14. Вавілов Є. В. Стан процесів управління виробництвом продукції та послуг на сучасному етапі у сенсі стандартів ISO 9000 / Є. В. Вавілов, Г. М. Котов // Технічне регулювання та якість: Сучасні стан та перспективи : V Всеукраїнська науково-практична конференція молодих учених, 20-21 травня 2014 р. : матер. конф. — Одеса, ОДАТРЯ. — С. 185-187.
15. Вавілов Є. В. Серія стандартів SQuaRE як основа забезпечення вимог до якості та оцінки програмних засобів / Є. В. Вавілов // Збірник наукових праць ОДАТРЯ. — 2015. — № 1(6). — С. 129-139.
16. Казакова Н. Ф. Автоматизація процесу адаптації інформаційних систем до інцидентів інформаційної безпеки / Н. Ф. Казакова, Є. В. Вавілов // Інформаційна безпека. — 2013. — №4(12). — С. 49-56.
17. ДСТУ ISO 9000:2007. Система управління якістю. Основні положення та словник. — Введ. 2008-01-01. — Київ : Держстандарт України, 2001. — 72 с.
18. ISO/IEC 25010:2011. Systems and software engineering – Systems and software Quality Requirements and Evaluation (SQuaRE) – System and software quality models [Електронний ресурс] : Портал : ISO. — Режим доступу \www/ URL: http://www.iso.org/iso/catalogue_detail.htm?csnumber=35733. — Заголовок з екрана, доступ вільний, 25.08.2014.
19. Грабовський О. В. Розробка методу оцінювання якості інформаційно-вимірювальних систем на основі використання генетичних алгоритмів : дис... канд. техн. наук: 05.01.02 / Грабовський Олег Вікторович; ОДАТРЯ. – Одеса, 2013. – 219 с.
20. Грабовський О. В. Регуляризація визначення показників якості функціонування ІВС з врахуванням нечіткості інформації / О. В. Грабовський, С. Л. Волков, О. О. Скопа // Вісник Національного технічного університету «ХПІ». — 2013. — №26 (999). — С.169-174.
21. Грабовський О. В. Показники якості та життєві цикли захищених інформаційно-вимірювальних систем / О. О. Скопа, С. Л. Волков, О. В. Грабовський // Вісник Східноукраїнського національного університету імені Володимира Даля. — 2013. — №15(204), ч. 1. — С. 192-198.
22. Грабовський О. В. Показники якості та життєві цикли інформаційно-вимірювальних систем / О. В. Грабовський, Т. І. Наконечна, С. Л. Волков // Збірник наукових праць Одеської державної академії технічного регулювання та якості. — 2012. — №1(1). — С. 17-23.

Дата надходження в редакцію: 14.08.2015 р.

Рецензент: д.т.н., проф. О. О. Скопа