UDC 519.7; 004.056

S. O. Kushch, *Ph.D.*
Cherkasy State Technological University
Schevchenko blvd, 460, Cherkasy, 18006, Ukraine
kushch@ieee.org

BINARY FORMS OF LOGIC FUNCTIONS REPRESENTATION

In the paper we attempt to perform an analysis and draw conclusions concerning the opportunities and perspectives of the use of new forms of Boolean functions representation. The aim of the paper is to consider new binary forms of Boolean functions representation – non-inverting classical algebraic and non-inverting classical Reed-Muller ones, which bring together so-called unary forms that exist today – algebraic and Reed-Muller representation forms – and do not contain defects of classical form of Boolean functions representation that is used in logical design. For the analysis we use "EDM – Extended Data Mining" technology, which is a generalization of the well-known "Data Mining" technology. Also major algorithms of Data Mining are provided. In addition, criteria and the technology of comparison of different forms of Boolean functions presentation, which is offered to use in this study, are presented. The conclusions concerning the prospects for further research are made.

Key words: *forms of Boolean functions representation, Boolean functions, binary Boolean functions, effectiveness of Boolean functions representation.*

Introduction. In the paper we attempt to perform an analysis and draw conclusions concerning the opportunities and perspectives of the use of new forms of Boolean functions representation.

As it has been considered in previous papers, the use of only classical representation of Boolean functions (CRF) makes sheet-oriented solutions non optimal in more than 90% of situations. The cause is that arguments x_i in classical representation of Boolean functions are used, as a rule, in two forms – direct and inverse. There is no such defect in algebraic (ARF) [1] and Reed-Muller (RMRF) [2] forms of Boolean functions representation. All these forms are called unary as all Boolean functions (BF) are implemented in one representation.

The optimal representation concept (ORC) lies in the fact that the representation, which gives the minimum value of selected index of structural complexity *S*, is used for any implementing logical functions (LF) and should be applied to "trinary" one.

It is obvious that except unary representations (CRF, ARF, RMFR) and trinary ones (TR) there are also so-called binary representations – presentation, in which conjunction addition element (CAE) provides the ability to add conjunctions in two forms. The development of binary representations (BR) allows to expect that advantages of different representations of Boolean functions will be combined and defects of CR will be compensated.

The paper is devoted to the investigation of binary representations (BR), in the first place to the study of their availability for using.

The selection of comparison test of Boolean functions

What are requirements for BR? First of all it is their completeness (the ability to implement any logical function in one representation). Concerning the problem of CRF default compensation, it is reasonable to include in BR only those LF in which x_i is used in one form (direct or inverse), in other words "non-inverting" LF. For completeness it is necessary to unite non-inverting LF with any full LF without inverting. It is possible to form two LF that meet this condition – **non-inverting classical algebraic and non-inverting classical Reed-Muller representations**.

- NICARF – non-inverting classical algebraic representation in which CAE enables both logical and algebraic (including weight numbers) conjunction adding;
- NICRMRF – non-inverting classical Reed-Muller representation in which CAE enables logical and Reed-Muller conjunction adding.

It is reasonable to investigate efficiency of both BR in comparison with the well-known representations (CRF, ARF, RMRF). For this purpose we use "EDM – Extended Data Mining" technology [3] that is a generalization of the well-known "Data Mining" technology.

Data Mining is the technology used for discovering of hidden correlations in large data sets. Two technologies are lying in the foundation of most of Data Mining tools: machine learning and visualization (image displaying). Imaging quality is determined by the capabilities of graphics display data. The graph representation variability with the aid of colors, shapes and other elements changing reduces the detection of hidden interactions.

Both technologies support each other in "Data Mining" analysis. The visualization is used to search exclusions, general trends and interactions and helps to select data at the initial stage of the project. Machine learning is used to search interactions in a debugged project.

Major algorithms of Data Mining are provided below.

Association rules explore cause-and-effect connections and determine the probability or validity coefficient, allowing to draw relevant conclusions. The rules are presented in the form of "if <condition> then <conclusion>". They can be used to predict or estimate unknown parameters (values).

Decision trees and subsumption algorithms determine natural "division" in the data, based on the goal variables. Firstly the division for the most important variables is performed. A tree branch can be represented as the adaptive part of the rule. The examples that can be found most often are algorithms of Classification and Regression trees (CART) or Chi-squared Automatic Induction (CHAID).

Sets of input signals, math functions enabling and value parameters weighting factor are used in artificial nerve nets to predict the target value. When the iterative search cycle is performed the nerve net is modifying the weighting factors until a predictive input value will correspond to the actual value. After such learning the nerve net becomes a model that can be applied to new data for prediction.

Genetic algorithm is a method that uses an iterative process of evolution sequence of model generation that includes reduction, mutation and junction operations. Fitness function is used for selection of intended persons and rejection of others. First of all genetic algorithms are used to optimize the topology of the nerve nets. However, they can be used independently, for modeling.

Memory-based Reasoning or Case-based Reasoning are algorithms based on the detection of some analogies in the past which are the closest to the current situation to assess an unknown value or to predict possible results (consequences).

Cluster analysis divides heterogeneous data on homogeneous or semi-homogeneous groups. The method allows to classify observations by a number of common features. The clusterization extends possibilities of forecasting.

Data Mining Technology is used in four main fields including science, business, research for government and Web-direction.

Data Mining is used to solve business problems. The main directions are Banking, Finance, Insurance, CRM, Production, Telecommunications, E-Commerce, Marketing, Stock Market and others.

Data Mining is used to solve problems at the state level. The main directions are the search of people who evade taxes and tools in the fight against terrorism.

Data Mining is used to solve Web issues. Main directions are search engines, counters and others.

Data Mining is used for scientific research. Main directions are medicine, biology, molecular genetics and genetic engineering, bioinformatics, astronomy, applied chemistry, research on drug dependence and others.

The expansion to Extended Data Mining Technology (EDM) [3] is that to solve scientific, technical, commercial or other issues we must form the so-called "EDM triplet " (Fig. 1) – a set of research object (RO), research methods (RM) and expected (forecasted) research results (ERR).

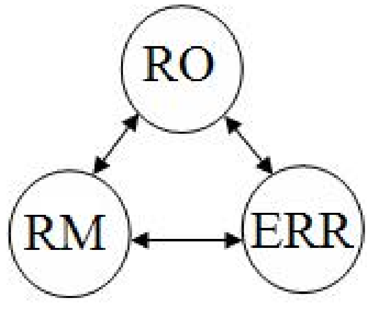

Fig. 1. EDM Triplet

The triplet elements of RO and RM in the research process can be formed in any sequence and they can change repeatedly during the life research cycle. The triplet can contain several close RO and several RM, each of them will produce different ERR. The comparison of obtained ERR and modification on this basis, the RM and RO for the next research cycle are distinctive feature of EDM. The main element of EDM triplet is RO that can be abstract and be a model of real object.

Going back to binary representation forms it should be noted that the alternatives availability (from fundamental scientific to current household) put a question of their comparison and the comparison requires developing technologies of some alternatives quantitative evaluation. This effectiveness comparison technology of different representation forms is offered in [3]. In this case this technology lets you to compare different representations forms. The area of programmable logic matrix (PLM) consists of two submatrixes. The conjunctions are formed in PLM1 submatrix, which are necessary for any form of LF representation. The formed conjunctions in PLM2 submatrix are added depending on selected RF, logical (LRF), algebraically (ARF) or through mod2 (RMRF).

To compare RF effectiveness integral indicators of the following form are offered:

- S_{AD} – the summands number in LF record, which identifies the entrances number of PLM2 submatrix;

- S_{SH} – the summands number in LF record that represents conjunction input arguments, which determines the lines number in PLM1 with sets of active elements;

- S_L – the letters number in LF record, which is classical criterion of LF minimization;

- S_S – dimensional square PLM, which is determined as:

$$S_S = 2nS_{AD} \text{ for CFR;}$$

$S_S = nS_{AD}$ for APF, RMRF and inverter less RF, where n is the number of input arguments PLM;

- S_{AC} – the area of active elements PLM, which is determined as:

$$S_{AC} = 2nS_{SH} \text{ for CFR;}$$

$$S_{AC} = nS_{SH} \text{ for APF and RMRF.}$$

It should be noted that from the point of view of concrete LF implementation specified values criteria can be considered as LF implementation indicators, which depend on selected PF.

Let's consider Fig. 2, in which for some selected S criterion graphs LF NLF are presented, which can be implemented in each PF for a given number of n arguments as a function of S criterion value.

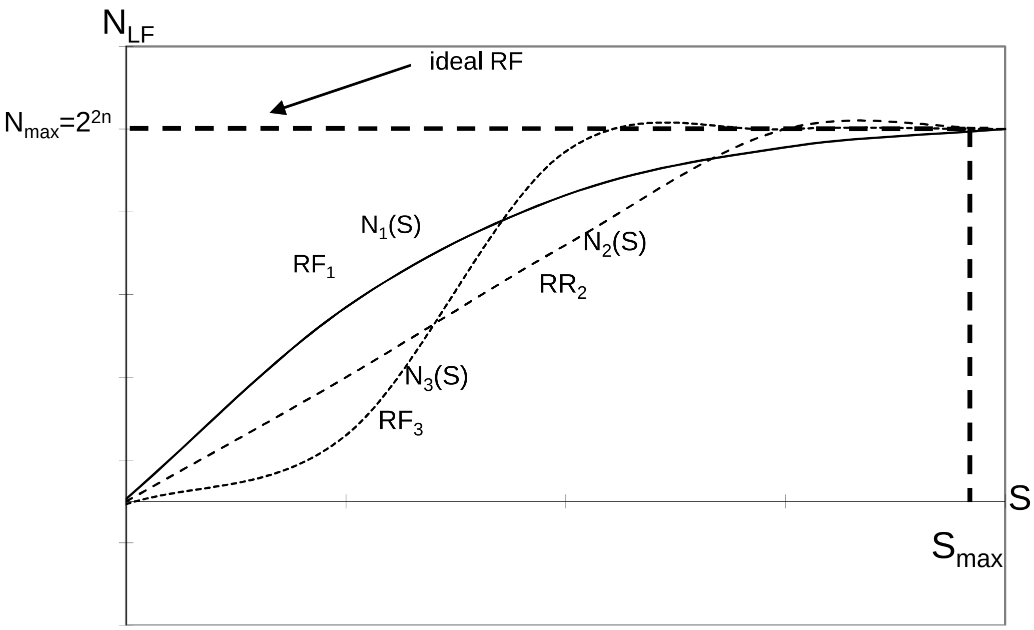

Fig. 2. The growth of LF NLF number that are implemented depending on *S* criterion value

Fig. 2 shows the graph NLF(S) for hypothetical ideal RF, which ensures the implementation of all LF from n arguments, the number of which are equal 22n by S = 0.

It is clear that every real RF as closer to the ideal as greater the curvilinear trapezoid area that are limited by Nj(S) curve on S [0, $S_{\max}$] changes interval.

So relative performance indicator (PRI) ***i***-th FP may be determined by the following formula (1).

$$h_i(n) = \frac{P_{i\ \text{real}}(n)}{P_{ideal}(m)} = \frac{\int_0^{S_{\max}(n)} N_j(S)ds}{N_{\max}(n) S_{\max}(n)}, \quad (1)$$

where $P_{i\ real}$ is an area above curvilinear trapezoid for the ***i***-th RF,

P_{ideal} is an curvilinear trapezoid area (in this case, the rectangle area) for the ideal RF.

All indicators S are integers, and also of piecewise-continuous nature of the functions $Ni(S)$, so the expression (1) can be written in the following equivalent form:

$$h_i = \frac{\sum_{j=0}^{S_{mm}} N_{ji}(S)}{N_{\max} S_{mm}} \quad (2)$$

where N_{ij} is the number of LF that are implemented on RLF, for a given value of the specified criterion,

$N_{\max}$ is the full LF amount of given n argument number,

S_{mm} is the maximum value of the selected criterion for all RF, ensuring the implementation of all LF [2].

If (2) can be rewritten in another form,

$$h_i = \frac{1}{S_{\max}} \sum_{j=0}^{S_{\max}} \frac{N_{ji}}{N_{\max}} = \frac{1}{S_{\max}} \sum_{j=0}^{S_{\max}} p_{ji}, \quad (3)$$

so that we can give a clear statistical content to the value of relative effectiveness indicator for any RF LF.

The formula (3) shows that p_{ji} is LF realization in the ***i***-th RF when the value of the selected criterion S <= j and the whole value of RPI h_i is the mean value of LF realization in the ***i***-th RF on the S selected criterion.

Conclusion. To obtain a clear answer concerning the use of the offered PF we need to conduct a great research in the following directions:

- to research classical algebraic forms efficiency of logic functions representation;

- to research the efficiency of classical Reed Muller representation form of logic function;

- to compare BRF1 and BRF2 effectiveness for complete sets of BF for different *n*.

The list of necessary research is not exhaustive and will be expanded during the work with non-inverting representations forms of Boolean functions. These Boolean functions form combinational circuits, which are information kernels of digital machine in modern systems of automatics, computer engineering, technical means of information protection and other elements of IT-technologies.

References

1. Kochkarev, Yu., Kushch, S. and Panasko, O. (2010) The analysis of consumer implementation indicators of Reed-Muller presentation form of logical functions. *Visnyk Cherkaskogo derzhavnogo tehnologichnogo universitetu*, (2), pp. 64-68.
2. Kochkarev, Yu., Kazarinova, N., Panteleeva, N. and Shakun, S. (1999) Reference book "Classical and alternative minimal forms of logical functions". Cherkasy, 195 p.
3. Kochkarev, Yu., Buzko, V. and Kucherova, N. (2006) Information technology EXTANDED DATA MINING for improving the structure of digital units and blocks. *Proceedings of the NSU*, (12). Dnepropetrovsk: NSU, pp.126-132.

Стаття надійшла до редакції 31.03.2014.

С. О. Кущ, *к.т.н., старший викладач кафедри інформатики та інформаційної безпеки*
Черкаський державний технологічний університет
б-р Шевченка, 460, м. Черкаси, 18006, Україна
Kushch@ieee.org

БІНАРНІ ФОРМИ ПРЕДСТАВЛЕННЯ ЛОГІЧНИХ ФУНКЦІЙ

Запропонована робота є спробою провести аналіз і зробити висновки щодо можливості та перспективності використання запропонованих нових форм представлення булевих функцій. Метою представленої роботи є розгляд нових форм представлення булевих функцій. В роботі пропонується розглянути нові бінарні форми представлення булевих функцій – безінверторну класико-алгебраїчну форму та безінверторну класико-Ріда-Мюллера форму, які об'єднують так звані унарні форми, що існують на сьогодні, – алгебраїчну та Ріда-Мюллера форми представлення – і не містять недоліків використовуваної при логічному проектуванні класичної форми представлення булевих функцій. Для аналізу доцільно використовувати технологію «EDM – Extended Data Mining», яка являє собою узагальнення відомої технології «Data Mining». Зроблені висновки щодо перспективності подальших досліджень.

Ключові слова: *форми представлення булевих функцій, булеві функції, бінарні булеві функції, ефективність представлення булевих функцій.*