

ОБРАБОТКА И ПЕРЕДАЧА ИНФОРМАЦИИ

УДК 681.391

Д. М. Романенко, Н. В. Пацей, А. М. Драко
Белорусский государственный технологический университет

ОСОБЕННОСТИ ДЕКОДИРОВАНИЯ ЛИНЕЙНЫХ БЛОЧНЫХ КОДОВ С ИСПОЛЬЗОВАНИЕМ РЕКУРРЕНТНЫХ И РАДИАЛЬНО-БАЗИСНЫХ НЕЙРОННЫХ СЕТЕЙ

В статье рассмотрены особенности применения нейронной сети для декодирования кода Хэмминга. Предлагается использовать нейронную сеть для декодирования полученного сообщения путем выбора к какому из вариантов принадлежит переданное сообщение, что является решением задачи классификации. Возможными вариантами для решения такой задачи являются рекуррентные нейронные сети и сети на основе радиально-базисных функций. Процесс декодирования рассматривается для кода Хэмминга (7, 4). Исходными данными при декодировании являются шестнадцать вариантов сочетания четырех информационных символов закодированного сообщения. Кодовые символы поступают на элементы сети, образующие входной слой. Данный слой использует рекуррентную связь между нейронами, что позволяет найти корреляцию между символами переданного сообщения. Выходной слой из шестнадцати элементов, связанный прямой связью с предыдущим слоем, выполняет функцию классификации, формируя тем самым биты декодированного сообщения. Сеть обучается на примерах входных и выходных значений элементов по методу градиентного спуска с минимизацией количества ошибок при классификации.

Выполнены исследования предложенных методов нейронного декодирования на основе разработанной программной модели. Согласно полученным результатам методы нейросетевого декодирования можно применять и для других блочных кодов, например, для итеративных. Нейросетевой декодер в целом показал высокие результаты по обнаружению и исправлению ошибок, которые ближе к «мягкому» декодированию, чем к «жесткому». Однако увеличение длины кодовой последовательности до 64 бит и выше существенно усложняет процесс декодирования.

Ключевые слова: кодирование, сети, нейрон, информация, каналы, связь, передача, обучение, уровни.

D. M. Romanenko, N. V. Patsei, A. M. Drako
Belarusian State Technological University

FEATURES OF LINEAR BLOCK CODES DECODING WITH USE OF RECURRENT AND RADIAL-BASIS NEURAL NETWORKS

The article describes the features of the decoding Hamming code with a neural network. When Hamming code (7, 4) is used, there are seven symbols that are transmitted over the network. However, only four symbols are informational which correspond to sixteen sets for messages after encoding. Possible variants for solving such a problem are made on base of recurrent neural networks and radially basic functions networks. The decoding process is considered. The initial data for decoding are sixteen variants of the four information symbols encoded message. This data build an input layer with recurrent connections between the elements. This layer will help to find a correlation between the symbols of the transmitted message. The output layer of sixteen elements, linked by a direct link to the previous layer, performs the classification function, thereby generating the bits of the decoded message. The network is trained on the examples of input and output values by the method of the gradient descent, that is based on the minimization of errors numbers during the classification.

Based on the developed software simulation model the proposed methods of neural network decoding are carried out. According to the obtained results, neural network decoding methods can be used with other block codes, for example, iterative. The neural network decoder as a whole showed high results in detecting and correcting errors, which are closer to “soft” than to “hard” decoding. However, increasing the length of the code sequence to 64 bits or more significantly complicates the decoding process.

Key words: coding, network, neuron, information channels, communication, connection, training, levels.

Введение. Рост числа пользователей коммуникационных систем, появление новых мультимедийных услуг предъявляют высокие требования к скорости, надежности и времени задержки при обработке информации. Одна из самых весомых задержек связана с необходимостью использования помехоустойчивого кодирования и декодирования данных. Для снижения временных затрат необходимо решать проблемы совершенствования существующих и разработки новых методов обработки информации.

В настоящее время подобные проблемы решаются различными путями, в том числе внедрением параллельных методов кодирования и декодирования, основанных на элементах теории искусственных нейронных сетей. Одним из перспективных направлений развития нейронных сетей является их использование при решении задач декодирования линейных блочных кодов в системах передачи информации, которые зачастую являются достаточно сложными для решения с использованием традиционной архитектуры вычислительных машин.

Существует множество подходов к использованию нейронных сетей как в качестве отдельных элементов системы декодирования, так и для построения дискретных декодеров [1]. В силу того, что декодер играет ключевую роль в процессе исправления ошибок в переданном сообщении, можно сделать вывод, что актуальность разработки новых методов декодирования, основанных на использовании нейронных сетей, не вызывает сомнений.

Основная часть. В целом нейросетевое декодирование можно рассматривать как процесс перехода от вторичного отображения сообщения (закодированного сообщения) к первичному алфавиту (информационному сообщению). Фактически оно сводится к задаче классификации или определения принадлежности входного образа, представленного вектором признаков, к одному или нескольким предварительно определенным классам. Сети с прямой связью являются универсальным средством аппроксимации функций, что позволяет их использовать в решении задач классификации. Как правило, нейронные сети оказываются наиболее эффективным способом классификации, потому что генерируют фактически большое число регрессионных моделей (которые используются в решении задач классификации статистическими методами).

К сожалению, в применении нейронных сетей в практических задачах возникает ряд проблем. Заранее неизвестно, какой сложности (размера) может потребоваться сеть для достаточно точной реализации отображения. Эта

сложность может оказаться чрезмерно высокой, что потребует сложной архитектуры сетей. Так, в работе [2] доказано, что простейшие однослойные нейронные сети способны решать только линейно разделимые задачи. Это ограничение преодолимо при использовании многослойных нейронных сетей. В общем виде можно сказать, что в сети с одним скрытым слоем вектор, соответствующий входной последовательности бит, преобразуется скрытым слоем в некоторое новое пространство, которое может иметь другую размерность, а затем гиперплоскости, соответствующие нейронам выходного слоя, разделяют его на классы. Таким образом, сеть распознает не только характеристики исходных данных, но и «характеристики характеристик», сформированные скрытым слоем. Последнее и позволяет найти корреляционные зависимости между информационными и избыточными символами сообщения.

Задача классификации при наличии двух классов может быть решена сетью из одного нейрона в выходном слое, который может принимать одно из двух значений (0 или 1) в зависимости от того, к какому классу принадлежит образец. При наличии нескольких классов возникает проблема, связанная с представлением этих данных для выхода сети. Наиболее простым способом представления выходных данных в таком случае является вектор, компоненты которого соответствуют различным номерам классов. При этом i -я компонента вектора соответствует j -му классу. Все остальные компоненты при этом устанавливаются в нуль. При интерпретации результата обычно считается, что номер класса определяется номером выхода сети, на котором появилось максимальное значение. Так, например, если в сети с тремя выходами имеется вектор выходных значений (0,3; 0,6; 0,4), то сообщение приравнивается ко второму классу, так как максимальное значение имеет вторая компонента вектора. При данном способе кодирования иногда вводится также понятие «уверенности» сети в том, что пример относится к этому классу. Наиболее простой способ определения «уверенности» заключается в установлении разности между максимальным значением выхода и значением другого выхода, которое является ближайшим к максимальному значению. Например, для рассмотренного выше примера «уверенность» сети в том, что пример относится ко второму классу, определится как разность между второй и третьей компонентой вектора и равна 0,2. Соответственно, чем выше уверенность, тем больше вероятность того, что сеть дала правильный ответ. Этот метод кодирования является самым простым, но не всегда оптимальным

способом представления данных (возникнут сложности при практической реализации при значениях $k > 64$).

Возможен также другой вариант классификации, который заключается в следующем: выходной вектор представляет собой номер кластера, записанный в двоичной форме. Тогда при наличии ста двадцати восьми классов потребуется вектор из восьми элементов, и, скажем, сто двадцать седьмому классу будет соответствовать вектор 01111111. При этом пятистам двенадцати классам соответствует вектор из десяти элементов, что всего на два элемента больше. В случае получения неверного значения на одном из выходов возможно получить неверную классификацию. В рассмотренном далее примере используется значение выходных элементов, равное количеству классов. Однако необходимо отметить, что обе описанные выше системы классификации имеют как преимущества, так и недостатки. Преимуществом первой будет надежность классификации, однако большое количество элементов сети делает ее сложной реализуемой и значительно более медленной, чем в случае, когда класс будет представлен двоичным числом.

При классификации большее значение имеет выбор данных. Так, для блочных кодов исходными данными являются информационная последовательность, которая подлежит преобразованию в кодовую. Поэтому можно использовать переданное сообщение как входные данные для нейросетевого декодера. Далее нейронная сеть при помощи скрытого слоя устанавливает корреляцию между информационными и избыточными битами, а на выходном слое будет определен класс, который укажет, какая информационная последовательность передавалась. При этом предполагается использовать обучение с «учителем». Входному сообщению будет соответствовать класс определенной кодовой последовательности.

Топология сети приведена на примере кода Хэмминга (7, 4). Кодовая последовательность подается на семь элементов на входном слое [3]. Скрытый слой также состоит из семи элементов, связанных рекуррентно, для получения корреляции между символами переданного сообщения. При этом информационных символов всего четыре, что соответственно дает шестнадцать вариаций кодовых последовательностей или классов для сети. В соответствии с этим используется либо шестнадцать элементов выходного слоя, либо четыре в зависимости от применяемого кодирования классов. Рассмотрим пример декодирования. Пусть после передачи по зашумленному каналу принято сообщение 0010110, которое и формирует входной

вектор нейронной сети. После выполнения соответствующих вычислений нейронной сетью будет получен выходной вектор значений: 1; 0,52; 1; 0,59; 1; 0,88; 0,62; 1; 0,86; 0,73; 1; 0,65; 1; 0,64; 0,59; 0,67. Минимальное значение, полученное во втором элементе вектора, означает, что переданная кодовая последовательность относится ко второму классу, который соответствует следующей кодовой последовательности 0010110 (согласно таблице соответствия класса кодовой последовательности). В итоге если выбранная нейронной сетью в процессе классификации последовательность совпадет с исходной информационной последовательностью, то можно считать, что декодирование завершено успешно.

Среди многослойных нейронных сетей можно выделить четыре наиболее подходящих варианта:

- сети прямого распространения;
- рекуррентные нейронные сети или сети обратного распространения (RNN);
- сети на основе радиально базисных функций (RFB);
- нейронная сеть Геленбе [4].

Отметим, что далее основное внимание будет уделено принципам декодирования с использованием двух из них: рекуррентных нейронных сетей и сетей на основе радиально-базисных функций.

Далее рассмотрим особенности применения рекуррентных нейронных сетей (Recurrent Neural Network, RNN). Основное отличие данных сетей от традиционных заключается в том, что каждый нейрон взаимодействует сам с собой. На вход данному типу сетей подается сигнал, который соответствует полученной кодовой последовательности. Каждый элемент такой последовательности поочередно передается одним и тем же нейронам, которые своё же предсказание возвращают себе вместе со следующим ее элементом до тех пор, пока последовательность не закончится. Элементы рекуррентной сети изображают как обычные нейроны с дополнительной циклической стрелкой, которая демонстрирует то, что кроме входного сигнала нейрон использует также свое дополнительное скрытое состояние. Если «развернуть» такое изображение, получится цепочка одинаковых нейронов, каждый из которых получает на вход свой элемент последовательности, выдает предсказание и передает его дальше по цепочке.

В соответствии с представленным выше описанием принципов декодирования с RNN рекуррентная нейронная сеть для декодирования кода Хэмминга (7, 4) будет выглядеть, как показано на рис. 1.

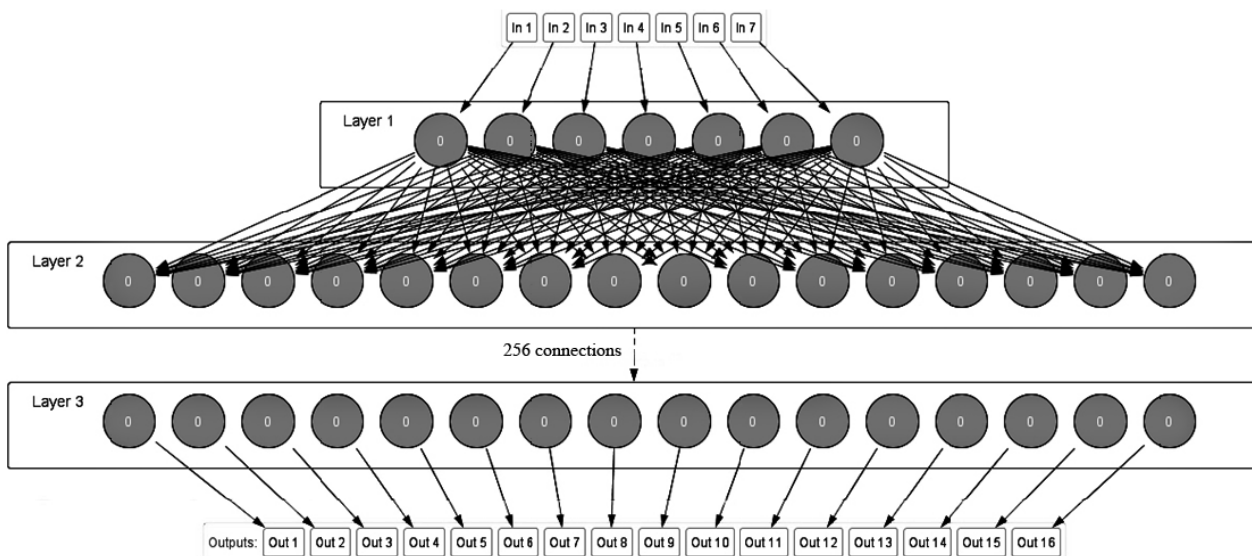


Рис. 1. Скриншот сгенерированной рекуррентной сети для декодирования кода Хэмминга (7, 4)

Другим типом исследуемых нейронных сетей являются сети, которые основываются на использовании радиально-симметричных функций. Это функции, относящиеся к специальному классу, характерное свойство которых заключается в том, что отклик функции монотонно убывает (возрастает) с удалением от центральной точки. Типичный пример такой функции – функция Гаусса, для скалярного аргумента имеющая следующее аналитическое представление:

$$h(x) = \exp\left(-\frac{(x - c)^2}{r^2}\right), \quad (1)$$

где c – смещение относительно центра; r – радиус (параметр, от значения которого зависит ширина функции).

В случае векторной переменной x формула для функции Гаусса приобретает следующий вид:

$$h(\bar{x}) = \exp\left(-\frac{\|\bar{x} - \bar{c}\|^2}{r^2}\right). \quad (2)$$

Структура однослойных RBF сетей, представлена на рис. 2.

Очевидно, что каждый j -й из n компонентов входного вектора подается на вход m базисных функций и их выходы линейно суммируются с весами:

$$\{w_j\}_j^m, \quad (3)$$

где w – вектор весов, показывающий смещение функции.

Таким образом, выход RBF сети является линейной комбинацией некоторого набора базисных функций:

$$f(\bar{x}) = \sum_{j=1}^m w_j h_j(\bar{x}). \quad (4)$$

В дальнейшем под функцией h будем подразумевать функцию Гаусса с векторным аргументом. Допускается использование в качестве базисных функций произвольный набор функций. В любом случае на основе некоторого набора экспериментальных данных типа «вход – выход» и в условиях полной неопределенности относительно вида возможной функциональной зависимости между входными и выходными данными «пытаются угадать» эту зависимость.

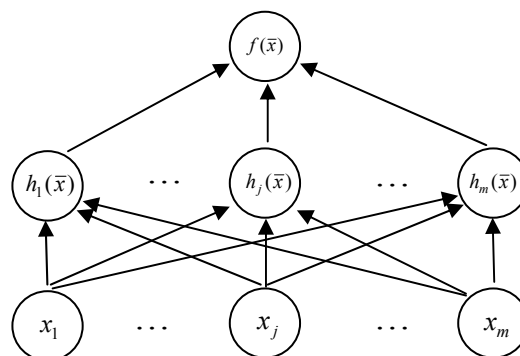


Рис. 2. Структура RBF сети для декодирования кода Хэмминга (7, 4)

Рассмотрим настройку RBF сетей. Предположим, что параметры функции h , а именно смещение c и радиус r , фиксированы, т. е. уже определены. Следовательно, задача нахождения весов в формуле (4) решается методами линейной алгебры. Этот метод называется методом псевдообратных матриц, он позволяет минимизировать средний квадрат ошибки. В соответствии с данным методом сначала находится интерполяционная матрица H :

$$H = \begin{pmatrix} h_1(\bar{x}_1) & h_2(\bar{x}_1) & \dots & h_m(\bar{x}_1) \\ h_1(\bar{x}_2) & h_2(\bar{x}_2) & \dots & h_m(\bar{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ h_1(\bar{x}_p) & h_2(\bar{x}_p) & \dots & h_m(\bar{x}_p) \end{pmatrix}. \quad (5)$$

В представленной матрице m – число нейронов в скрытом слое; p – размер обучающей выборки, а каждый вектор x состоит из n компонентов, где n – число входов сети.

На следующем этапе вычисляется инверсия произведения матрицы H на транспонированную матрицу H^T :

$$A^{-1} = (H^T H)^{-1}. \quad (6)$$

Окончательный результат в виде вектора весов рассчитывается по формуле

$$\bar{W} = A^{-1} H^T \bar{y}. \quad (7)$$

Если предыдущее предположение о фиксированных параметрах функции активации h не выполняется, т. е. помимо весов необходимо настроить параметры активационной функции каждого нейрона (смещение функции и ее радиуса), задача становится нелинейной. И решать ее приходится с использованием итеративных численных методов оптимизации, например, градиентных методов.

Для изучения процесса декодирования с использованием нейронных сетей и сравнения с классическими методами декодирования была разработана программная модель, имитирующая работу следующих типов декодеров: классического «жесткого», «мягкого», нейронного на основе рекуррентной сети, нейронного на основе RBF сети. В результате моделирования процесса декодирования были получены результаты декодирования, представленные в виде зависимостей вероятности ошибки от уровня шума (рис. 3 и 4).

В ходе моделирования процесса декодирования было установлено, что использование нейронного декодера на основе RBF сети уступает по эффективности «мягкому» декодированию, однако лучше, чем «жесткое» декодирование и несколько превосходит декодирование на основе простой рекуррентной сети, которая в свою очередь также превосходит декодеры жесткого типа.

При этом необходимо отметить, что нейронные сети могут не только декодировать переданное кодовое сообщение, но и также давать оценку успешности декодирования, что может быть важным при появлении в процессе передачи многократных ошибок высокой кратности, превосходящих корректирующие возможности используемого помехоустойчивого кода.

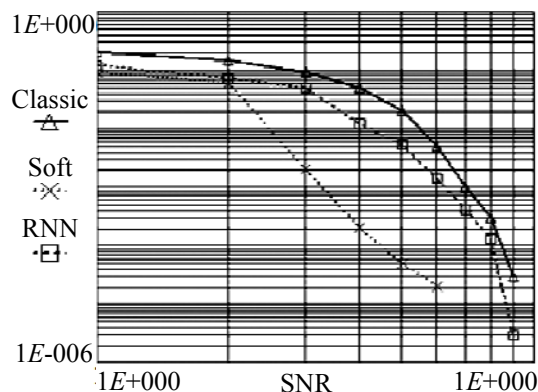


Рис. 3. Зависимости вероятности ошибки после декодирования от уровня шума для классического декодера, «мягкого» и декодеров на основе рекуррентной сети

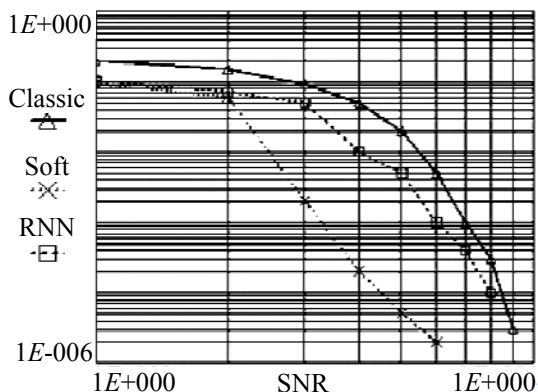


Рис. 4. Зависимости вероятности ошибки после декодирования от уровня шума для классического декодера, «мягкого» и декодеров на основе RBF сети

Несмотря на то что эффективность «мягкого» декодера в целом не была достигнута, необходимо отметить достаточно большой потенциал методов декодирования, основанных на использовании нейросетевых технологий, который заключается в возможности выполнения параллельных операций (например, вычислений) при аппаратном исполнении декодера, например, на основе VLSI (Very-large-scale integration) чипа. Также необходимо отметить, что при классическом методе декодирования кода Хэмминга возможно исправление только одной ошибки, а при декодировании с использованием нейронных сетей возможно исправление до трех ошибок с относительно высокой вероятностью получения на выходе декодера корректного информационного сообщения.

Заключение. По результатам проведенных исследований можно утверждать, что нейронные сети можно использовать для декодирования линейных блочных кодов, например, кода Хэмминга. В процессе декодирования нейрон-

ная сеть выполняет задачу классификации. Предложенные методы декодирования на основе рекуррентной и RBF нейронных сетей предполагают универсальность применения к всем кодам, относящимся к классу линейных блочных, и показывают высокие результаты по обнаружению и исправлению ошибок, которые по эффективности хоть и уступают «мягкому» декодированию, но при этом превосходят «жесткое». Однако увеличение длины кодовой последовательности значительно усложняет процесс декодирования. Выходом из данной си-

туации может служить кодирование классов в виде последовательности бит, что значительно уменьшает количество выходных элементов нейронной сети либо использование многослойных нейронных сетей. Разработанная имитационная модель процесса декодирования линейных избыточных кодов с использованием нейросетевых технологий позволила получить оценку эффективности предложенных методов на основе рекуррентной и RBF нейронных сетей в виде зависимостей вероятности появления ошибки от соотношения сигнал/шум.

Литература

1. Zeng G., Hush D., Ahmed N. An application of neural net in decoding error-correcting codes // *IEEE International Symposium on Circuits and Systems*. 1989. P. 782–785.
2. Minsky M. L., Papert S. *Perceptrons: An Introduction to Computational Geometry*. Cambridge, MA: MIT Press, 1969. 258 p.
3. Stefano A. D., Cataldo G. D., Palumbo G. On the use of neural networks for hamming coding // *IEEE International Symposium on Circuits and Systems*. 1991. Vol. 3. P. 1601–1604.
4. Драко А. М., Романенко Д. М. Особенности нейросетевого декодирования линейных блочных кодов // *Труды БГТУ*. 2015. № 6: Физ.-мат. науки и информатика. С. 166–170.

References

1. Zeng G., Hush D., Ahmed N. An application of neural net in decoding error-correcting codes. *IEEE International Symposium on Circuits and Systems*. 1989, pp. 782–785.
2. Minsky M. L., Papert S. *Perceptrons: An Introduction to Computational Geometry*. Cambridge, MA: MIT Press, 1969. 258 p.
3. Stefano A. D., Cataldo G. D., Palumbo G. On the use of neural networks for hamming coding. *IEEE International Symposium on Circuits and Systems*, 1991, vol. 3, pp. 1601–1604.
4. Drako A. M., Romanenko D. M. Features of neural network decoding of linear block codes. *Trudy BGTU* [Proceedings of BSTU], 2015, no. 6: Physical-mathematical sciences and informatics, pp. 166–170 (In Russian).

Информация об авторах

Романенко Дмитрий Михайлович – кандидат технических наук, заведующий кафедрой информатики и веб-дизайна. Белорусский государственный технологический университет (220006, г. Минск, ул. Свердлова, 13а, Республика Беларусь). E-mail: rdm@belstu.by

Пацей Наталья Владимировна – кандидат технических наук, исполняющая обязанности заведующего кафедрой программной инженерии. Белорусский государственный технологический университет (220006, г. Минск, ул. Свердлова, 13а, Республика Беларусь). E-mail: n.patsei@belstu.by

Драко Арсений Михайлович – магистр технических наук, ассистент кафедры информатики и компьютерной графики. Белорусский государственный технологический университет (220006, г. Минск, ул. Свердлова, 13а, Республика Беларусь). E-mail: a.drako@belstu.by

Information about the authors

Romanenko Dmitri Mikhailovich – PhD (engineering), Head of the Department of Informatics and Web-design. Belarusian State Technological University (13a, Sverdlova str., 220006, Minsk, Republic of Belarus). E-mail: rdm@belstu.by

Patsei Nataliya Vladimirovna – PhD (engineering), Acting Head of the Department of Programming Engineering. Belarusian State Technological University (13a, Sverdlova str., 220006, Minsk, Republic of Belarus). E-mail: n.patsei@belstu.by

Drako Arseni Mikhailovich – Master of Engineering, Assistant, the Department of Informatics and Computer Graphics. Belarusian State Technological University (13a, Sverdlova str., 220006, Minsk, Republic of Belarus). E-mail: a.drako@belstu.by

Поступила 28.11.2017