Нейросетевая кластеризация узлов беспроводной сенсорной сети

Представлен анализ возможностей кластеризации узлов беспроводной сенсорной сети (БСС) с помощью нейронной сети — самоорганизующейся карты Кохонена (СКК). В ходе исследования построена модель БСС, которая основана на таких параметрах как координаты узлов, либо уровень сигнала между ними. В перспективе возможно использовать также уровень остаточной энергии. Синтезирована модель СКК с параметрами, адекватными для кластеризации узлов БСС. Для произведения экспериментов, модели реализованы в виде программного обеспечения, позволяющего располагать узлы БСС в двухмерном пространстве с учетом радиуса мощности их приемопередатчиков, а также применять для кластеризации СКК. В результате выявлено, что СКК эффективна для кластеризации узлов БСС. При этом ее эффективность зависит от формы представления входных данных. Кластеризация узлов БСС с помощью СКК сети является эффективным направлением, поскольку при кластеризации можно задать множество входных параметров, а СКК в свою очередь позволяет проецировать многомерные входные данные на двумерную решетку — карту Кохонена. Последняя, информативно описывает распределение кластеров БСС в пространстве. СКК показала свою эффективность в случае кластеризации по матрице мощностей.

Ключевые слова: беспроводная сенсорная сеть, нейронная сеть, самоорганизующаяся карта Кохонена, кластеризация, моделирование.

Maxpoв C.C., аспирант МТУСИ, SlavaM4@vandex.ru

Введение

В результате проведенного анализа, представленного в работе [1], можно сделать вывод, что многоинтервальная иерархическая модель связности узлов является наиболее эффективной для построения топологии БСС. Данная модель позволяет увеличить время жизни всей сети благодаря оптимальной маршрутизации пакетов между узлами.

Согласно этой модели, вся сеть делится на кластеры, во главе каждого из которых выбирается главный кластерный узел (ГКУ), на который передаются данные с остальных узлов кластера. Далее, агрегированные данные ГКУ передают на базовую станцию (БС) напрямую или опосредованно через соседние ГКУ, используемые в качестве ретрансляторов.

Кластеризация узлов, в свою очередь, является задачей, от успешного решения которой зависит эффективность работы всей сети. Если узлы будут неправильно кластеризованы, то они не смогут передавать данные в рамках своего кластера, поскольку другие узлы будут вне досягаемости мощности приемопередатчиков. Соответственно, необходим механизм, который позволит производить кластеризацию узлов оптимальным образом.

Обязательным условием для кластеризации является наличие критериев, в соответствии с которыми она будет производиться. Для кластеризации узлов БСС возможно использовать следующие критерии:

- уровень мощности узлов по отношению к остальным узлам;
 - географические или относительные координаты;
 - уровень остаточной энергии.

Следующей немаловажной задачей, без которой невозможна кластеризация — это определение количества кластеров, которое для одних и тех же данных может быть совершенно разным. Проблема определения числа кластеров является одной из основных нерешенных до настоящего времени задач кластерного анализа [2]. Существуют лишь формальные подходы, при этом, чтобы эффективно определить количество кластеров, необходимо учитывать специфику предметной области, в рамках которой необходимо кластеризовать элементы.

Среди таких формальных подходов можно выделить ряд алгоритмов: Gap-подход, G-Means и X-Means. Но данные алгоритмы предполагают повторное выполнение кластеризации с последующей оценкой качества полученной кластерной структуры [3]. Повторное выполнение кластеризации производится, поскольку, указанные алгоритмы сами являются алгоритмами кластеризации, которые выполняют разведывательную функцию — экспериментальное выделение кластеров и вместе с тем, соответственно, определение количества кластеров. Далее, после применения таких алгоритмов, производится уже окончательная кластеризация с помощью алгоритмов кластеризации. В результате, применение таких методов требует значительных вычислительных затрат для решения этой задачи.

Таким образом, кластеризация в предложенной модели топологии БСС является основным структурным элементом для произведения самоорганизации сети.

Одним из эффективных инструментов кластеризации является нейронная сеть. Нейронные сети обладают рядом преимуществ:

- массовый параллелизм;
- распределённое представление информации и вычисления:
 - способность к обучению и обобщению;
 - алаптивность:
 - свойство контекстуальной обработки информации;
 - устойчивость к ошибкам.

T-Comm #6-2014 31

ТЕХНОЛОГИИ

Для кластеризации БСС была выбрана нейронная сеть Кохонена, которая позволяет представить многомерные входные данные, характеризующие беспроводные узлы, в виде двумерной или трехмерной карты Кохонена.

Цель и методы исследования

Целью данной работы является исследование возможности самоорганизации БСС посредством нейронной сети – самоорганизующейся карты Кохонена (СКК).

Для достижения цели были поставлены следующие задачи:

- Изучение математического аппарата кластеризации СКК;
- Выявление эффективных критериев кластеризации БСС;
 - Интеграция БСС с СКК;
- Произведение компьютерного моделирования кластеризации БСС с помощью СКК.

Описание математического аппарата

Основой математического аппарата нейронной сети Кохонена, как и других нейронных сетей, является подстройка весов нейронов, путем вычисления Евклидова расстояния (возможен выбор также другой метрики) между векторами весов нейросети и входным вектором. В нейросети Кохонена в отличие от других используется соревновательный механизм обучения и после прохождения входного вектора через сеть, на её выходе ненулевой уровень сигнала имеет лишь один выходной нейрон — нейрон-победитель. Этот нейрон определяет кластер, к которому в результате, относится входной вектор.

Одну из ключевых ролей в математическом аппарате сети Кохонена играет функция соседства и коэффициент скорости обучения. Функция соседства (neighborhood function) позволяет при подстройке весов нейрона-победителя влиять также на веса соседних нейронов в некоторой окрестности вектора. Данная окрестность, которую образует вокруг себя нейрон-победитель, называется радиусом обучения (learning radius). При этом, радиус обучения максимален на первой итерации и уменьшается по мере увеличения числа итераций. В результате в конце обучения на і-ом векторе, происходит подстройка только весов нейрона победителя. Коэффициент скорости обучения позволяет управлять скоростью обучения сети Кохонена. На начальных итерациях, когда веса нейронов далеки от оптимальных, скорость обучения достаточно высока, в результате чего происходит грубая подстройка весов. По мере увеличения количества итераций, скорость обучения замедляется, а подстройка весов становится точнее [6]. Более подробно математический аппарат описан в алгоритме обучения, который будет описан далее в данной статье.

Определение эффективных критериев кластеризации

В ходе произведенных опытов, путем компьютерного моделирования, было выявлено, что эффективнее всего производить кластеризацию БСС с помощью СКК по такому критерию как географические координаты узлов, поскольку они наиболее полно позволяют представить вектора узлов. СКК в свою очередь, как было описано ранее, производит кластеризацию на основании отличия (расстояния) векторов друг от друга. При наличии координат х, у (возможно использования также z), она эффективно может выделить кластеры в БСС. С практической точки зрения использование в качестве критерия географических координат узлов неэффективно, в виду следующих обстоятельств:

- повышение себестоимости узла сети за счет GPSмодуля;
- при определенных условиях сигнал может не доходить до приемника, или приходить со значительными искажениями или задержками;
- зависимость условий получения сигнала от министерства обороны США [7].

Соответственно, более эффективным критерием было бы использование уровней мощности узлов по отношению к другим узлам, поскольку он имеет ряд значительных пре-имуществ:

- не требует аппаратной интеграции в БСС;
- не зависит от связи со спутниками;
- позволяет узнать фактическое наличие канала и уровня мощности этого канала между узлами.

Таким образом, в качестве входных данных можно использовать уровни мощности сигналов узлов по отношению друг к другу, или другими словами — энергетической видимости узлов. Данные о видимости узлов можно наглядно представить в виде неориентированного графа, где в процентах будет отображена мощность между всеми узлами сети. Пример такого графа представлен на рис. 1.

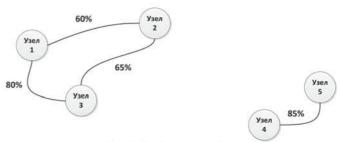


Рис. 1. Граф мощностей

Запишем данный граф в виде матрицы мощностей (видимостей) узлов:

$$P = \begin{pmatrix} \infty & 60 & 80 & 0 & 0 \\ 60 & \infty & 65 & 0 & 0 \\ 80 & 65 & \infty & 0 & 0 \\ 0 & 0 & 0 & \infty & 85 \\ 0 & 0 & 0 & 85 & \infty \end{pmatrix}$$

В результате, мы представили каждый из 5 узлов БСС в виде 5-мерного вектора, компонентами которого являются уровни видимости узлов.

Кластеризация матрицы мощностей невыполнима с помощью алгоритмов кластеризации графов (например, алгоритм кластеризации Маркова / Markov's clustering algorithm / MCL), поскольку они ориентированы на классическое представления графа в виде матрицы расстояний. Здесь же имеет место быть неопределенность в расстоянии между узлами, обозначенная в матрице символом бесконечности ∞.

Подготовка входных данных и инициализация сети Кохонена

Является необходимым произвести нормализацию векторов для того, чтобы с ним корректно могла оперировать нейронная сеть. Нормализацию, соответственно будем производить по формуле:

32 T-Comm #6-2014

$$p_{NORM} = \frac{p}{p_{MAX}}$$
, где $p_{MAX} = 100$.

На примере предыдущей матрицы, получим:

$$P = \begin{pmatrix} 1 & 0.6 & 0.8 & 0 & 0 \\ 0.6 & 1 & 0.65 & 0 & 0 \\ 0.8 & 0.65 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0.85 \\ 0 & 0 & 0 & 0.85 & 1 \end{pmatrix}$$

Определим количество кластеров, на которые будет разделена сеть на основании полученной матрицы.

Вычислим среднюю плотность распределения мощности сети, то есть число узлов, которое в среднем видит каждый узел сети по следующей формуле:

$$\rho_{CPEJI} = \frac{\sum_{i=1}^{N} \rho_i}{N},$$

где ρ — число видимых узлов для i-го узла, N — количество узлов сети.

Число кластеров найдем по формуле:

$$K = \frac{N}{\rho_{CPE/I}}$$

Зададим параметры сети Кохонена:

- Количество входных нейронов N будет равняться количеству узлов БСС;
- Количество нейронов слоя Кохонена (выходных нейронов) будет равняться количеству кластеров K, которое мы определили ранее;
- Инициализируем веса нейронов случайными числами в диапазоне от 0,0 до 1,0;
- Для каждого нейрона введем такой параметр как потенциал.

Для обучения сети по алгоритму Кохонена характерна проблема "мертвых" нейронов, которая состоит в том, что одним из ограничений конкурирующего слоя - слоя Кохонена является то, что ряд нейронов остается незадействованными. Нейроны никогда не выигрывают в конкурирующем слое, поскольку их вектора значительно отличаются от векторов обучения, что инвариантно к длительности обучения. В результате, такие нейроны никогда не становятся победителями. Это приводит к тому, что входные данные будут интерпретироваться меньшим числом классов и нейронов, соответственно. Погрешность квантования при этом возрастает. Следовательно, таким "мертвым" нейронам необходимо дать шанс выиграть. Для этого в алгоритм обучения добавляют такой параметр для каждого нейрона как потенциал. При инициализации сети нейронам присваивается потенциал $p_i^{t=0} = \frac{1}{n}$, где n - число нейронов кластеров, где t — ну-

левая итерация обучения (этап инициализации).

При этом, вводится также понятие минимального потенциала p_{\min} . Если значение потенциала i-го нейрона слоя Кохонена p_i становится меньше минимального p_{\min} , то данный нейрон Кохонена исключается из рассмотрения — "отдыхает". Для значений, которыми возможно инициализировать p_{\min} характерны следующие закономерности:

• при $p_{\min} = 0$ нейроны не исключаются из борьбы;

• при $p_{\min} = 1$ нейроны побеждают по очереди, так как в каждом цикле обучения только один из них готов к борьбе.

На практике хороший результат получается при $p_{\min} \approx 0.75$ [6].

Соответственно, проинициализируем потенциал p_i для каждого нейрона слоя Кохонена равным $p_i \approx 0.75$. Далее, на каждой итерации в будем пересчитывать потенциалы нейронов слоя Кохонена согласно формуле (2), которая будет приведена далее в статье.

Подготовка обучающей выборки

Подготовим обучающую выборку, с помощью которой обучим сеть Кохонена выделять кластеры в данных об узлах сети. Каждый узел сети описывает вектор видимости. Из матрицы P стохастически выберем количество L векторов обучающей выборки согласно правилу, которое было экспериментально выведено в результате исследований и изучения работы нейросети Кохонена при выделении кластеров в ECC:

если
$$2K \ll N$$
, то $L = 2K$, иначе $L = K$

где K — количество кластеров, N — число узлов (входных нейронов), L — размер обучающей выборки.

Алгоритм обучения нейронной сети Кохонена

Рассмотрим алгоритм обучения сети Кохонена согласно [4, 5] с настроенными коэффициентами функций, используемый для кластеризации узлов БСС по матрице нормализованным значениям векторов матрицы мощностей.

Обучение состоит из двух циклов:

- внешнего на каждой итерации происходит исполнение внутреннего цикла по всем обучающим векторам;
- внутреннего происходит перебор каждого обучающего вектора $E_1...E_L$.

Внешний цикл — это основной цикл, которые будет продолжаться до тех пор, пока веса нейронов изменяются. Будем считать, что изменения весов нейронов прекратились, как только веса нейронов предыдущей итерации основного цикла равны весам текущей итерации основного цикла с точностью $\varepsilon = 0,0001$.

Внутренний цикл по всем обучающим векторам Е1...ЕL:

1) Определяем нейрон-победитель для рассматриваемого на текущей операции обучающего вектора. Нейрономпобедителем будет тот нейрон, вектор весов которого в наименьшей степени отличается от вектора весов обучающего вектора. То есть, расстояние между их векторами весов должно быть минимально. В качестве метрики используем Евклидово расстояние, которое вычисляем по формуле (1):

$$d(x, w_i) = ||x - w_i|| = \sqrt{\sum_{j=1}^{n} (x_j - w_{ij})^2},$$
(1)

2) После определения нейрона победителя, подстраиваем потенциалы нейронов по формуле (2):

$$p_{i}^{t+1} = \begin{cases} p_{i}^{t} + \frac{1}{n}, & i \neq j, \\ p_{i}^{t} - p_{\min}, & i = j, \end{cases}$$
 где p_{i}^{t} – потенциал i -го нейрона, j – номер нейрона победи-

где p_i^t – потенциал i-го нейрона, j – номер нейрона победителя, t – номер итерации обучения, n – число выходных нейронов (нейронов слоя Кохонена).

3) Подстраиваем веса (компоненты вектора) нейронапобедителя и близлежащих нейронов по формуле (3):

$$w_i^{t+1} = w_i^t + h(t) \cdot \alpha(t) \cdot (x - w_i^t), \tag{3}$$

где w_i^t – вектор весов і-го нейрона, t – номер итерации обучения, x – входной вектор, h(t) – функция соседства, $\alpha(t)$ – коэффициент скорости обучения.

На данном этапе происходит подстройка весов всех нейронов Кохонена. Величина изменения весов w_i^k каждого нейрона зависит от произведения разности каждого веса w_i^k и компонента входного вектора х на коэффициент скорости обучения $\alpha(t)$ и некоторый коэффициент усиления, которым является функция соседства h(t).

Для сходимости необходимо, чтобы выполнялось условие [4]: $\lim h(t) = 0$.

В качестве функции соседства зададим функцию Гаусса (4):
$$h(t) = \exp\left(-\frac{d(w_i, w_j)}{2\sigma^2(t)}\right), \tag{4}$$

где $d(w_i, w_i)$ – Евклидово расстояние между *i*-м нейроном w_i и нейроном-победителем w_i , σ является убывающей функцией от номера цикла обучения.

Согласно [4] σ и $\alpha(t)$ будем вычислять по формулам (5):

$$\sigma(t) = \sigma_0 \cdot \exp\left(-\frac{t}{\tau}\right) \quad \text{if } \alpha(t) = \alpha_0 \cdot \exp\left(-\frac{t}{\tau}\right) \tag{5}$$

В результате произведенных экспериментов, выявлено, что эффективнее всего задать коэффициент $\sigma_0 = \alpha_0 = 0.1$, а временную константу т для обоих формул задать равной $\tau = 1000$. Тогда, получим:

$$\sigma(t) = \alpha(t) = 0.1 \cdot \exp\left(-\frac{t}{1000}\right).$$

Компьютерное моделирование кластеризации БСС посредством СКК

Для произведения экспериментов в области кластеризации, была разработана моделирующая среда [8], которая позволяет наносить на двухмерную поверхность беспроводные узлы, рассчитывать матрицу мощностей и применять для кластеризации СКК.

Каждый узел имеет радиус действия своего приемопередатчики и отображен в интерфейсе зеленым кругом (рис. 2).

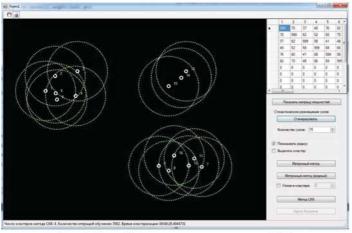


Рис. 2. Интерфейс среды моделирования БСС

В правом верхнем углу интерфейса отображена матрица мощностей для текущего расположения узлов на 2D-поверхности. Матрица мощностей, не смотря на плоское двухмерное расположение узлов БСС, является адекватной моделью, соответствующей реальности, так как вне зависимости от числа измерений, матрица отображает видимость узлов друг от относительно друга, принимая во внимание лишь величину уровня сигнала.

Произведем кластеризацию 15 узлов с помощью СКК, в результате на рабочей области приложения радиусы узлов будут закрашены в разные цвета, соответствующие класте-

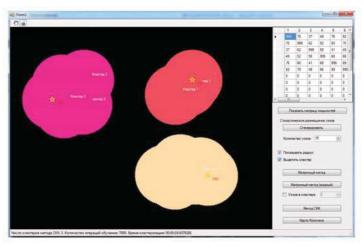


Рис. 3. Результат кластеризации СКК

Увеличим количество узлов до 10000. Получим следующее отображение в интерфейсе программного обеспечения (рис. 3).



Рис.4. Результат кластеризации 10 000 СКК

Время, затрачиваемое на кластеризацию 5-20 узлов, составляет от 1 секунды до минуты в зависимости от распределения случайных весов нейронов при инициализации СКК. Количество итераций колеблется от 3000 до 10000.

На кластеризацию 10 000 узлов БСС уходит около 4-5 ч. Кластеризация производилась с указанной ранее точностью $-\varepsilon = 0.0001$. При уменьшении точности, скорость кластеризации может значительно возрасти. Количество итераций составляет около 10 000.

T-Comm #6-2014 34

Заключение

Известно, что нейронная сеть позволяет эффективно производить кластеризацию объектов на основании такого критерия как географические координаты. Но географические координаты не всегда могут дать точное представление о реальной связности узлов – их энергетической видимости. Следовательно, эффективно использовать такой критерий кластеризация как энергетическая видимость, на основании которого в качестве аналитического представления в данной работе составлялась матричное представление – матрица мощностей.

Понятие матрицы мощностей схоже с понятием матрицы расстояний. Большинство алгоритмов кластеризации не подходят для кластеризации матрицы расстояний. Это объясняется тем, что в случае с графом, необходимо учитывать вершины графа и все связи. В то же время существует не так много алгоритмов, позволяющих производить кластеризацию по графу. Одним из наиболее известных является алгоритм кластеризации Маркова. Поэтому большинство алгоритмов кластеризации не подходят для данной задачи.

В случае использования матрицы мощностей, алгоритмы кластеризации графов уже не подходят для этой задачи, поскольку в данном типе аналитического представление используется иное представление графа, где имеет место неопределенность расстояния между отдельными его вершинами.

В результате проведенных исследований, экспериментальным путем было выяснено, что нейронная сеть - СКК является эффективным методом кластеризации матрицы мощностей, которая в свою очередь отражает фактическую энергетическую видимость всех узлов сети по отношению друг к другу. В большинстве случаев были получены корректные результаты кластеризации, которые становились качественнее по мере того, плотность узлов возрастала. Для более точной кластеризации, следует продолжить исследования в направлении подстройки параметров функции соседства СКК, а также пост-обработки нормализованных данных, которые подаются на вход СКК.

Одно важное явление, касающееся формирования обучающей выборки также необходимо отметить. Согласно изначальной версии алгоритма Кохонена [5], обучающая выборка может формироваться из реальных данных, путем случайного выбора некоторого числа обучающих векторов. Но при таком стохастичном подходе в обучающую выборку могут не попасть данные из некоторых кластеров, которые существуют. Следовательно, эффективнее обучать нейронную сеть на всем наборе реальных данных об узлах сети.

Литература

- 1. Махров С.С. Исследование связности узлов в иерархических протоколах беспроводных сенсорных сетей. Фундаментальные проблемы ра-диоэлектронного приборостроения / Материалы Международной научно-технической конференции «INTERMATIC-2013», 2-6 декабря 2013 г., Москва / Под ред . акаде-мика РАН А.С. Сигова. М.: Энергоатомиздат, 2013, часть 4. С.118-123.
- 2. Яцкив И., Гусарова Л. Методы определение количества кластеров при классификации без обучения. Transport and Communication. Vol.4, №1, 2003. C.23-28.
- 3. *Болдак А.А.*, *Сухарев Д.Л*. Определение количества кластеров в статистических данных. Вісник НТУУ «КПІ» Інформатика, управління та обчислювальна техніка №53. С.118-122.
- 4. Баталов А.С. Методы повышения эффективности обучения нейронной сети Кохонена //. Вестник Пермского университета. Сер.: Математика. Механика. Информатика, 2012. Вып. 3(11). С.86-93.
- 5. Кохонен Т. Самоорганизующиеся карты. Пер. 3-го англ. изд. М.: БИНОМ. Лаборатория знаний, 2008. 655 с.: ISBN 978-5-94774-352-4.
- 6. Горбаченко В.И. Сети и карты Кохонена: [Электронный ресурс] // Научно-исследовательский центр самоорганизации и развития систем. М., 2010-2014. URL: http://gorbachenko.selforganization.ru/index.html (Дата обращения: 01.02.2014).
- 7. GPS (Global Positioning System): [Электронный ресурс] // Санкт-Петербургский техникум геодезии и картографии. М., 2006-2014. URL: http://www.spbtgik.ru/book/6220.htm (Дата обращения: 01 02 2014)
- 8. *Шелухин О.И., Тенякшев А.М., Осин А.В.* Моделирование информационных систем. М.: Радиотехника, 2005. 368 с.

Neural clustering of wireless sensor network nodes

Mahrov S.S., MTUCI, Moscow., SlavaM4@yandex.ru

Abstract. In article presented an analysis of the possibilities of clustering nodes of wireless sensor network (WSN) by neural network - Kohonen's self-organizing map (SOM). During research was created a model of WSN, which is based on such parameters as the coordinates of nodes or a signal level between them and a level of residual energy. Also synthesized SOM model with parameters valid for clustering WSN's nodes. For the producing an experiments, models are developed as software (modeler), which allows locate WSN's node in 2D space with the power radius of transceiver, as well as apply for clustering SOM. As a result, it was found that the SOM is effective for clustering WSN's nodes. Moreover, its performance depends on the form of the input data representation. WSN's nodes clustering by SOM is an effective direction, because as clusterization criteria is possible to set a many parameters, and the SOM allows projection a multidimensional input data to a two-dimensional lattice — Kohonen's map. The latter describes a informatively describes a WSN's clusters in space. In addition, the SOM has proven its effectiveness in the case of clustering the capacity matrix.

Keywords: wireless sensor network, neural network, Kohonen's self-organizing map, clustering, modeling.

References

- 1. Mahrov S.S. Study connected nodes in a hierarchical protocol for wireless sensor networks. Fundamental problems of pa-dioelektronnogo Instrumentation / Proceedings of the International Scientific and Technical Conference «INTERMATIC-2013", 2-6 December 2013, Moscow, 2013, part 4. pp.118-123.
- 2. Yatskiv I.L. Methods Gusarova quantification of clusters in the classification without training. Transport and Communication. Vol.4, No 1, 2003. pp. 23-28.
- 3. Boldak A.A., Sukharev D.L. Determination of the number of clusters in a statistical data. News of NTU "KPI" No53. pp.118-122.
- 4. Batalov A.S. Methods to improve the effectiveness of training Kohonen neural network / Bulletin of the University of Perm. Ser.: Mathematics. Mechanic. Computer Science, 2012. Vol. 3 (11). pp. 86-93.
- 5. Kohonen T. self-organizing maps. Lane. 3rd Eng. ed. Moscow, 2008. 655 p.: ISBN 978-5-94774-352-4.
- 6. Gorbachenko VI. Network and the Kohonen map [Electronic resource] / Research Center and self-development systems. Moscow, 2010-2014. URL: http://gorbachenko.self-organization.ru/index.html (Access Date: 01.02.2014).
- 7. GPS (Global Positioning System): [electronic resource] / URL: http://www.spbtgik.ru/book/6220.htm (Access Date: 01.02.2014).
- 8. Shelukhin O.I., Tenyakshev A.M., Osin A.V. Modeling of information systems. Moscow, 2005. 368 p.

T-Comm #6-2014 35