Analytical and numerical modeling of the process for cluster emergence of objects in a random environment

Y A Mostovoy\textsuperscript{1} and V A Berdnikov\textsuperscript{1}

\textsuperscript{1}Samara National Research University, Moskovskoe Shosse 34A, Samara, Russia, 443086

e-mail: teridax63@yandex.ru, jakob.mostovoi@yandex.ru

Abstract. The task solution for probing a certain area by a swarm of objects randomly distributed is considered. This requires studying of statistical features for clusters emergence from swarm objects uniformly distributed in order to find such a concentration of the objects in the sensing area in which the number of random clusters will be maximal. Corresponding analytical dependencies have been obtained.

1. Introduction

One of the most rapidly developing areas of robotics today is the group robotics of mobile objects (or swarm robotics) \cite{1, 2, 3, 4, 5, 6}. The swarm robotics is a system of interconnected robots – in comparison with a conventional single robot having a number of advantages, the most significant of which are:

- high speed and range of such a system due to distribution of the swarm throughout the service area;
- high probability for accomplishing the assigned task due to possibility for redistribution of targets and replacement of the failed robot with another from the swarm;
- variety of options for achieving the system objectives by redistributing roles between the swarm objects;
- simplicity of the tasks solved by each robot while interacting with a lot of the same swarm objects allows solving rather complicated computational tasks.

Many tasks of robotics for mobile objects come to the percolation theory tasks in particular to the search for optimal concentration of objects to realize two-phase operations \cite{7, 8, 9, 10}.

The classical percolation theory considers a matrix with random filling as a model of a random operating environment in a direct geometric interpretation \cite{11, 12, 13, 14, 15, 16}. In this square matrix with the number of lines, the \( L \) random cell part is "black" conducting the liquid or gas flow, transport or information flow, and the remaining cells are "white" not conducting the flow. While increasing the concentration (probability of appearance) of black cells, some of them randomly begin to touch one another with edges, which may be interpreted as formation of close interaction, and merge. The "black" cells touching the edges form random conducting clusters that form and increase together with increase in concentration of "black" cells \cite{11, 17}.

In the classical percolation theory \cite{11, 15, 16, 17, 18}, concentration of "black" cells \( p_{\text{th}} \) is sought – the threshold of stochastic percolation, at which a through random route is formed on the "black" cells through the entire matrix in the intended direction – the stochastic percolation cluster. However, the stochastic percolation cluster has a loose structure, many "dead branches" and is clearly redundant for purposes of solving many practical tasks.

While implementing programmable percolation \cite{7, 8, 9, 10} in the first phase, a stochastic base is created from randomly distributed objects at values of their concentration much lower than the
threshold of stochastic percolation, and in the second phase, a percolation route is built by virtue of the introduction (installation) of additional objects at available intercluster intervals. In this case, concentration of the stochastic base is chosen in such a way that the total costs for a two-phase operation are minimal.

The specific dependence of the average number of clusters detected as a result of numerical statistical modeling on the concentration with the maximum value at the concentration in the area is0.25 physically explainable: With further increase in concentration, the clusters increase and begin to combine actively while their number decreases.

Since the concentration of the objects, at which the number of the clusters has a maximum, is important for the mentioned cost-effective two-phase operations in a number of applied tasks, analytical determination for exact value of this statistical phenomenon for cluster formation on matrices with random filling is needed.

2. Task statement
Swarm robotics is considered. Each robot can interact with a neighboring one in the process of solving some element of the global problem. The maximum number of clusters in the swarm of the objects allows them to be easily switched and combined to solve tasks jointly at the right place of the service area. This place can be assigned efficiently [7, 8, 9, 10, 19].

Let us consider the following task. At what concentration (probability for location of the object in the cell) of the objects on the percolation matrix with the number of cells will the number of the clusters be maximal? For the swarm its size $S = Np_{\text{max}}$, being $N$ the number of cells of percolation matrix ($N = l_1 l_2$), $p_{\text{max}}$ – the concentration at which the number of the clusters is maximal.

In order to answer this question, let us introduce a special function $M(p)$ that will consider the dependence of the average number of clusters on concentration $p$.

3. An analytical task solution for finding the maximum number of the swarm clusters
Let us consider a cluster consisting of one element. Such cluster can exist only if one percolation matrix cell appears black with probability, and the adjacent ones are white [11, 17] to ensure independence of the cluster and to keep it from merging during "planting" on the matrix (creating a stochastic base [7, 8, 9, 10, 11]). In other words, the similar cluster can be represented as follows: $p(1 - p)^4$. For a cluster of two cells, similar reasoning will "return" the following form: $p^2(1 - p)^6$. A cluster of three cells will have the form $p^3(1 - p)^7, or p^3(1 - p)^8$, because it can be either in the form of a "corner" (the number of empty neighbors is 7), or a "line" (the number of empty neighbors is 8). Continuing to increase the number of cells in the cluster, we inductively obtain that for a certain number of cells, the cluster form will be appeared as follows [11, 17]:

$$p^s(1 - p)^t$$

being $p$ – concentration, $s$ – number of the cells which the cluster consists of, $t$ – required number of empty (white) cells around the cluster.

Let us consider the possible structures, types and number of clusters that arise on the matrix with random filling. A cluster of two cells can be either a vertical "line" or a horizontal one. A cluster of three cells in the form of a "corner" on the matrix can be located in four positions, the case with the cluster of three cells in the form of a "line" is analogous to a cluster of two cells. In other words, the average number of the clusters of three cells is an expression of the form $4p^3(1 - p)^7 + 2p^3(1 - p)^8$[17]. Developing these arguments (presented in the work of Y.Y. Taraskevich [17]), we obtain a formula for the average number of clusters on the matrix of infinite size:

$$\sum_{t=t_{\text{min}}}^{t_{\text{max}}} g_{st} h^s(1 - p)^t$$

being $t_{\text{min}}$ – the minimum number of empty neighboring cells, $t_{\text{max}}$ – the maximum number of empty neighboring cells, $g_{st}$ – the number of different clusters that can be obtained from the cells and empty neighboring cells.
\[ t_{\text{min}} = \left\lceil 4\sqrt{s} \right\rceil \quad (3) \]

\[ t_{\text{max}} = 2s + 2 \quad (4) \]

As the number of cells increases, the geometric length and/or width of the cluster will approach the matrix length and/or width, respectively. Someday, a moment will come when one of these parameters coincides. Then the subsequent increase in the number of the clusters for the length and/or width will not be possible on the area matrix \((N = l_1 l_2)\), in other words, must be limited by parameter \(g_{st}\).

Then the average number of clusters from these elements on the matrix with the number of cells \(N\) will be:

\[ \sum_{t=t_{\text{min}}}^{t_{\text{max}}} g_{st} np^s (1-p)^t \quad (5) \]

Total number of clusters on the matrix with the number of cells \(N\):

\[ M(p) = \sum_{s=1}^{s=N} \sum_{t=t_{\text{min}}}^{t_{\text{max}}} g_{st} np^s (1-p)^t \quad (6) \]

being \(M(p)\) – required function of the average number of clusters.

Then concentration of the maximum number of clusters is the concentration at which the maximum value of the function appears \(M(p)\).

Let us assume the function \(M(p)\) in the following form:

\[ M(p) = \sum_{s=1}^{s=N} M_s(p) \quad (7) \]

In view of the large number of coefficients \(M_s(p)\) and absence of an analytic expression for the parameter \(g_{stN}\) (this parameter can only be determined by searching all possible clusters with fixed parameters \(s\) and \(t\), which is a complicated computational task), we try to approximate this function by another one, easier to study.

Also, because of the knowledge of the limited number of parameters, if \(g_{stN}\) is of interest to investigate the ratios of the coefficients \(M_s(p)\) relative to each other. To do this, let us fix any \(p\) from 0 to 1, for example \(p = p^*\), and look at the behavior of the sums \(M_s(p)\), namely: try to determine which\(s\) contributes the most to the resulting sum \(M(p)\).

The result is that \(M_{s=k}(p) > M_{s=k+1}(p)\) and in \(M(p)\), just the first few components make major contribution. The appearance of the components \(M_s(p)\) is shown in Figure 1.

It is also easy to determine that \(\int_0^1 M_{s=k}(p)dp > \int_0^1 M_{s=k+1}(p)dp\). For this, we calculate for each \(pM_{s=k}(p)\) and we \(M_{s=k+1}(p)\) sum them obtaining integrals of \(M_{s=k}(p)\) and \(M_{s=k+1}(p)\) integrals in the Lebesgue form [20, 21].

\[ \forall p \quad M_{s=k}(p) > M_{s=k+1}(p) \Rightarrow \sum_{i=p_{\text{min}}}^{p_{\text{max}}} M_{s=k}(p) > \sum_{i=p_{\text{min}}}^{p_{\text{max}}} M_{s=k+1}(p) \Rightarrow \int_0^1 M_{s=k}(p)dp \geq \int_0^1 M_{s=k+1}(p)dp \quad (8) \]

Having counted the first six integrals, we note that at \(s = 6\) \(\int_0^1 M_{s=6}(p)dp = o(p)\) i.e. is a small quantity and does not make a significant contribution to \(M(p)\).
Figure 1. Appearance of functions $M_s(p)$: schedule 1 – $M_{s=1}(p)$, schedule 2 – $M_{s=2}(p)$, schedule 3 – $M_{s=3}(p)$, schedule 4 – $M_{s=4}(p)$, schedule 5 – $M_{s=5}(p)$.

Then:

$$M(p) \approx \sum_{s=1}^{5} \sum_{t=t_{\text{min}}^{s}}^{t_{\text{max}}^{s}} g_{stN} p^s (1-p)^t$$

(9)

Such expression is well approximated by the modified function, the Rayleigh distribution density [22] (Figure 2).

Figure 2. Approximation of the function by the $M(p)$ modified Rayleigh distribution density function. Schedule 1 is the modified Rayleigh distribution density function, schedule 2 is the function $M(p)$, schedule 3 is the function $M(p)$ obtained in the course of the mathematical experiment.

Analyzing the schedules in Figure 2, we note that the maximum points on the schedules coincide, and the maxima are just at the concentration 0.26.

Schedule 1 on this figure is the Rayleigh distribution density function modified by introducing the parameter $C$ expressed as follows:

$$f(p, C, \sigma) = C \frac{p}{\sigma^2} e^{-\frac{p^2}{2\sigma^2}}$$

(10)

being $\sigma$ – Rayleigh function parameter, $p$ – concentration of the objects – probability for availability of the object in the cell, $C$ – normalization coefficient modifying the Rayleigh function and numerically equal to the definite integral of $M(p)$:
\[ C \approx \int_{0}^{1} M(p) \, dp \]  

(11)

being \( p \) – concentration, \( M(p) = \sum_{s=1}^{N} \sum_{t=t_{\text{min}}}^{t_{\text{max}}} g_{stN} p^s (1 - p)^t \).

Investigation of the functions \( f(p, C, \sigma) \) and \( M(p) \) has shown that the points of their maxima coincide and are equal \( p_{\text{max}} = 0.26 \). In other words, at this concentration, the maximum number of clusters on the percolation matrix is observed, which means that the number of local tasks solved by a swarm of size \( S = N p_{\text{max}} \), and hence also of robot clusters will also be maximal.

4. Statistical modeling for formation of clusters from swarm objects

An extensive mathematical experiment has been conducted to study formation of clusters from swarm objects with their uniform distribution ("planting") along the matrix. The experiment has been carried out as follows: The objects with different concentrations have been planted on many percolation matrices (Figure 3).

![Figure 3. Example of the planted percolation matrices of different concentrations (0.25; 0.4).](image)

Further, by the Hoshen-Kopelman algorithm [23, 24], the clusters, the number of which has been considered and normalized depending on the matrix size, have been distinguished (the experiment has been carried out for matrices of size \( 50 \times 50, 100 \times 100 \) and \( 200 \times 200 \)). The result, normalized for the matrix area, is shown in Figure 4 [7, 8, 9, 10, 19].

The result obtained has been approximated by the method of least squares with 200 points modified by the Rayleigh function \( f(p, C, \sigma) \) (Figure 2) with respect to the parameters \( C \) and \( \sigma \). The normal equations for the least squares method are as follows:

\[
\begin{align*}
C &= -\sum_{i=1}^{200} y_i \left( p_i \frac{p_i^2}{\sigma^2} \right) - \frac{1}{\sigma^2} \sum_{i=1}^{200} y_i \left( p_i \frac{p_i^2}{\sigma^2} \right) \\
&\quad + \frac{1}{\sigma^2} \sum_{i=1}^{200} y_i \left( p_i \frac{p_i^2}{\sigma^2} \right) \\
&\quad + \frac{1}{\sigma^2} \sum_{i=1}^{200} y_i \left( p_i \frac{p_i^2}{\sigma^2} \right)
\end{align*}
\]  

(12)

being \( y_i \) – the data obtained in the course of the mathematical experiment \( p_i \) – concentration, \( C \) – normalizing factor, \( \sigma \) – Rayleigh law parameter. From the second equation, \( \sigma \) can be determined by the numerical procedure.

The values of the parameters \( C \) and \( \sigma \) obtained as a result of approximation of the function \( M(p) \) by the normalized Rayleigh function \( f(p, C, \sigma) \) are equal, respectively, 0.0496 and 0.26 that coincides with the value of the definite integral of the function \( M(p) \) for the first five components and the value of the number of degrees of freedom for the function \( f(p, C, \sigma) \).
In other words, the results of statistical modeling, mathematical analytical modeling and their approximation by the modified Rayleigh distribution function have coincided with satisfactory accuracy.

**Figure 4.** Outcome of the statistical analysis.

5. Conclusions

As a result of the work done, the following may be concluded:

1. The maximum number of clusters in the swarm of the objects allows them to be easily switched and combined to solve tasks jointly at the right place of the service area. This place can be assigned efficiently \([7, 8, 9, 10, 19]\). The concentration of the swarm objects on the service area surface, at which the number of robot clusters is maximal, is equal to \(p_{\text{max}} = 0.26\).

2. The concentration value obtained by analytical modeling is confirmed by statistical mathematical modeling. The swarm size is expressed as \(S = Np_{\text{max}}\), being \(N\) – the number of the cells in the percolation matrix \((N = l_1l_2)\), \(S\) – the swarm size, \(p_{\text{max}}\) – concentration.

3. The obtained form of the curve is well approximated by the modified Rayleigh function \(f(p, C, \sigma) = C \frac{p}{\sigma^2} e^{-\frac{p^2}{2\sigma^2}}\).

6. References

[1] Kalyaev I A, Gayduk A R and Kapustyan S G 2009 Models and algorithms of team management in robot teams (Moscow: Fizmatlit) p 279

[2] Kalyaev I A and Gaiduk A R 2004 Social principles of management in the entity groups Mechatronics, Automation, Control 27-38

[3] Abraham A, Grosan C and Ramos V 2006 Swarm Intelligence in Data Mining Studies in Computational Intelligence 20

[4] Beni G 2005 From swarm intelligence to swarm robotics Lecture Notes in Computer Science 3342 1-10

[5] Bonabeau E and Meyer Ch 2001 Swarm intelligence: a whole new way to think about business Harvard Business Review 79 106-114

[6] Sahin E 2005 Swarm robotics: From sources of inspiration to domains of application Lecture Notes in Computer Science 3342 10-20

[7] Mostovoy Y A 2011 Statistical phenomena for scale distributed clusters of nanosatellites Bulletin of the Samara National Research University 26 80-89

[8] Mostovoy Y A 2013 Two-phase operations in scale networks of nanosatellites Computer Optics 37 120-130

[9] Mostovoy Y A 2013 Controlled percolation and optimal two-phase operations in scale networks of nanosatellites Bulletin of the Samara National Research University 11 53-62
[10] Mostovoy Y A 2015 Modeling of optimal two-phase operations in random operating environments Optoelectronics, Instrumentation and Data Processing 51 35-41
[11] Moskalev P V and Shitov V V 2007 Mathematical modeling of porous structures (Moscow: Fizmatlit) p 120
[12] Shen A 2004 Mathematical induction MCCME 36
[13] Alexandrowicz Z 1980 Critically branched chains and percolation clusters Physics Letters A 80 284-286
[14] Sarshar N, Boykin P O and Roychowdhury V P 2004 Scalable percolation search in power law networks Proceedings of the Fourth International Conference on Peer-to-Peer Computing 2-9
[15] Stauffer D 1979 Scaling theory of percolation clusters Physics Reports 54 1-74
[16] Stauffer and Aharony A 1992 Introduction to Percolation Theory (London: Taylor & Francis) p 180
[17] Tarasevich Y Y 2002 Percolation: theory, applications, algorithms (Moscow: URSS) p 109
[18] Galam S and Mauger A 1996 Universal formulas for percolation thresholds Phys. Rev. E 53(3) 2177-2181
[19] Mostovoy Y A and Berdnikov VA 2017 Large scale networks security strategy CEUR Workshop Proceedings 1901 187-193
[20] Kolmogorov A N and Fomin SV 1976 Elements for the theory of functions and functional analysis (Moscow: Science) p 544
[21] Trenogin V A 1980 Functional Analysis (Moscow: Science) p 495
[22] Perov A I 2003 Statistical theory of radio engineering systems (Moscow: Radio Engineering) p 400
[23] Babalievski F 1998 Cluster counting: Hoshen-Kopelman algorithm vs. Spanning three approach International Journal of Modern Physics 43-61
[24] Hoshen J and Kopelman R 1976 Percolation and cluster distribution. I. Cluster multiple labeling technique and critical concentration algorithm Phys. Rev. B 14 3438-3445