

유전알고리즘을 이용하여 최적화된 방제 자원 배치안의 분포도 분석

김혜진¹, 김용혁^{2*}

¹광운대학교 컴퓨터과학과 석사, ²광운대학교 컴퓨터과학과 교수

Distribution Analysis of Optimal Equipment Assignment Using a Genetic Algorithm

Hye-Jin Kim¹, Yong-Hyuk Kim^{2*}

¹Master, Dept. Computer Science, Kwangwoon University

²Professor, Dept. Computer Science, Kwangwoon University

요약 해양 오염사고를 대비한 계획으로, 최적화된 배치안들을 수집하여 분석하는 연구가 필수적이지만, 해양 오염사고 대응을 위한 최적을 배치안을 다양화하고 분석한 연구는 아직 선행되지 않았다. 이러한 필요성에 따라, 우리는 방제 자원 배치 최적화를 위한 유전알고리즘을 고안하고 이를 통해 최적의 방제 자원 배치안을 10,000 개 도출하였다. k -평균 알고리즘으로 군집화한 결과, 예상 최대 유출지역인 여수, 대산, 울산에 대하여 두 개의 군집으로 확연히 구분되었다. 우리는 이러한 군집을 새몬 맵핑을 통해 이차원으로 사영하여 배치안의 분포도를 분석하였고, 군집에 포함되는 배치안들이 그렇지 않은 배치안보다 시뮬레이션의 결과가 우수함을 확인했다. 향후, 본 연구를 기반으로 성능이 우수한 근사 모델을 구현하는 것이 가능할 것으로 보인다.

주제어 : 융합, 유전알고리즘, 클러스터링, 새몬맵핑, 자원배치최적화

Abstract As a plan for oil spill accidents, research to collect and analyze optimal equipment assignments is essential. However, studies that have diversified and analyzed the optimal equipment assignments for responding to oil spill accidents have not been preceded. In response to the need for analyzing optimal equipment assignments study, we devised a genetic algorithm for optimal equipment assignments. The designed genetic algorithm yielded 10,000 optimal equipment assignments. We clustered using the k -means algorithm. As a result, the two clusters of Yeosu, Daesan, and Ulsan, which are expected to be the largest spills, were clearly identified. We also projected 16-dimensional data in two dimensions via Sammon's mapping. The projected data were analyzed for distribution. We confirmed that results of the simulation were better than those of optimal equipment assignments included in the cluster. In the future, it will be possible to implement an approximate model with excellent performance based on this study.

Key Words : Convergence, Genetic algorithm, Clustering, Sammon's mapping, Optimal resource allocation

*This research was a part of the project titled 'Marine Oil Spill Risk Assessment and Development of Response Support System through Big Data Analysis', funded by the Korea Coast Guard.

*Corresponding Author : Yong-Hyuk Kim(yhdfly@kw.ac.kr)

Received January 28, 2020

Accepted April 20, 2020

Revised March 31, 2020

Published April 28, 2020

1. 서론

방제자원배치는 대표적인 실세계의 자원배치 최적화(resource allocation optimization) 문제이다. 이는 제한된 자원을 상황에 맞게 할당 하는 방법을 효율적으로 찾는 문제이며, 수자원[1], 제조[2]와 같은 다양한 분야로의 활용가치가 높은 할당 문제이다. 그러나 자원배치 최적화문제는 높은 차원의 조합최적화 문제들이 대부분이며, 이를 해결하기 위해 해공간을 모두 탐색하여 글로벌 최적해를 찾는 것은 현실적으로 불가능하다. 이러한 이유로 자원배치 최적화를 해결하기 위해 적합도에 기반한 휴리스틱(heuristic) 탐색[3]을 통해 최적해를 찾는다. 자원배치 최적화 문제에는 다양한 상황과 제약을 갖는 문제들이 속하므로, 휴리스틱 탐색보다 유연하게 해를 찾아내는 메타 휴리스틱(meta heuristic)[4]을 사용하는 것이 적절하다.

이처럼 복잡한 실세계의 자원배치 문제를 해결하기 위해, 대표적인 메타 휴리스틱 기법인 유전알고리즘을 통해 접근한 연구들이 있다. Dai와 Wang[5]은 그리드 서비스 신뢰도를 기준으로 평가하는 그리드 서비스 할당문제에 대한 최적화 모델을 제시했다. 이를 유전알고리즘을 통해 해결하였으며, 상관 장애가 없을 경우 거의 최적의 해를 찾을 수 있었다. Lavric 등[6]은 오염물질과 저수준의 자원을 갖춘 수자원 시스템을 최적화하기 위해 유전알고리즘으로 접근하였다. 입력과 출력 장치의 제약 조건과 함께 오염정도에 따라 평가되며, 수학적으로 프로그래밍한 기법과 비교하여 유전알고리즘을 이용한 기법의 우수함을 확인했다. Guerrero 등[7]은 클라우드 아키텍처의 컨테이너의 자원 할당을 위해 유전알고리즘을 이용한 접근법을 제안했다. 해당 연구에서는 클라우드 클러스터, 컨테이너, 마이크로서비스를 포함한 네 가지의 최적화 목표로 알고리즘이 동작한다. 유전알고리즘으로 최적화된 자원을 기반으로 시스템 프로비저닝, 시스템 성능, 시스템 장애 및 네트워크 오버헤드를 향상시켰으며, 쿠버네티스(Kubernetes)에서 구현된 컨테이너 관리 정보보다 컨테이너 할당 및 탄력성 문제를 해결하기에 적합함을 보였다.

본 논문의 구성은 다음과 같다. 2 절에서는 방제자원 최적화 문제에 대하여 기술하고, 3 절에서는 방제자원배치 최적화를 위해 고안된 유전알고리즘에 대하여 설명한다. 4 절에서는 k -평균 클러스터링을 통해 군집들에 대하여 분석하고, 5 절에서는 새몬뎀핑을 이용하여 16 차원의 배치안 데이터를 2 차원으로 사영하여 분포도를 분

석한다. 마지막으로, 6 절에서 결론을 내리며 본 논문을 마무리한다.

2. 방제자원배치 최적화

현실의 제약조건과 지역별 적정량 소요량을 모두 반영하는 실험을 구성하기 위해 Yoon 등[8]이 작성한 국가 방제능력 산정방안 연구의 최종보고서를 기반으로 진행하였다. 해당 보고서에서 방제자원배치는 해양 오염 발생에 대응하기 위해, 유출된 기름을 회수하는 장비인 유회수기를 적절히 한국의 16 개 지역으로 분산 배치 최적화한다. 또한, 유출된 기름의 1/3을 유회수기를 이용하여 해상 회수하기 위한 최적의 배치를 찾는 것이 목표이다. 방제자원배치는 대표적인 실세계의 자원 배치 문제이기 때문에, 많은 제약조건이 존재한다. 본 논문에서는 기름 유출 시나리오를 우리나라의 각 지역에서 최대로 유출될 수 있는 기름의 양으로 한다(Table 1 참고).

Table 1. Maximum oil spill accident scenario

| Location | Oil spill accident (kl) |
|--------------|-------------------------|
| Incheon | 8,500 |
| Pyeongtaek | 1,200 |
| Deasan | 45,000 |
| Gunsan | 3,800 |
| Mokpo | 8,500 |
| Wando | 600 |
| Yeosu | 45,000 |
| Jeju | 800 |
| Seogwipo | 500 |
| Tongyeong | 1,700 |
| Changwon | 1,200 |
| Busan | 2,500 |
| Ulsan | 45,000 |
| Pohang | 800 |
| Donghae | 500 |
| Sokcho | 50 |
| Total | 165,650 |

그러나 뚜렷한 평가방법이 없는 실세계의 문제에 대하여 해를 도출하고 평가하며 최적화하는 것은 매우 어려운 일이다. 본 연구에서는 해당 문제에 대한 평가를 위해, 실제 기름 유출사고 상황과 유사한 시뮬레이션을 고안하였다. 최대유출량의 목표량을 회수하는 시간을 최소화하기 위한 시뮬레이션을 수식화하면 다음과 같다.

함수 $f(x, i)$ 는 배치된 유회수기로 지역 i 에 유출량을 회수하는데 걸리는 작업시간이다. 따라서, 본 시뮬레이션을 이용한 유전알고리즘의 목적함수는 다음과 같이 표현할 수 있다.

$$\min \sum_{i=1}^n f(\mathbf{x}, i) \quad (1)$$

i 지역의 배치된 회수용량과 최대 유출량을 각각 x_i , Q_i 라 한다. 여기서 \mathbf{x} 는 각 지역의 배치된 회수용량을 나타내며, 다음과 같이 표현한다.

$$\mathbf{x} = (x_1, x_2, \dots, x_n), \quad x_i \geq 0, \quad x_i \in \mathbb{N} \quad (2)$$

$$Q_i \in \mathbb{R}$$

t_{ij} 는 지역 i 에서 지역 j 로 동원되는데 걸리는 소요시간으로 출항준비시간, 항해시간을 의미한다. 지역 i 에서 사고가 발생했을 때, t_{ij} 시간 이후가 작업개시일이고, 3일 이내에 8시간씩 작업해서 모두 회수해야 한다. 즉, 작업시간 기준으로 24시간 이내에 회수해야 하는 것이 목표이다. 08:00~18:00 사이에 작업을 하는 것이 일반적인 상황이다. 함수 $h(t)$ 는 동원하는데 걸리는 시간 t 를 작업개시가 12시일 때, 동원하는 동안 작업 못하는 시간으로 변환하는 함수이다. 예를 들어 사건이 작업개시가 오전 12시에 일어났고, 지역 i 로 동원하는데 걸리는 시간이 9시간인 경우, 그동안 작업을 못하는 시간인 $h(9)$ 는 1이다. 해당 함수는 다음과 같이 표현할 수 있다.

$$\begin{aligned} h(t) &= h(t-24) + T_w, \\ h(t) &= \min(T_w, t - T_{w_0}) \\ \text{if } T_{w_0} &\leq t \leq T_{w_0} + 24 \end{aligned} \quad (3)$$

함수 $g(t_1, t_2)$ 는 작업시간이 t_1 이고 작업개시 시간이 t_2 이고 인 경우, 동원 동안에 작업하지 못하는 시간이다. 이는 다음과 같이 표현할 수 있다.

$$\begin{aligned} g(t_1, t_2) &= h(t_1 + t_2) - h(t_2), \\ t_{ij} &\geq 0, \quad t_{ij} \in \mathbb{N} \end{aligned} \quad (4)$$

예를 들어 작업개시가 9시에 일어나고 동원시간이 3시간인 경우, $h(3+9) - h(9)$ 가 된다. $h(12)$ 는 12시 작업개시로부터 12시간 후이므로 4이고, $h(9) = 1$ 이므로, 3시간 동안 작업을 못한 것이 된다.

실험의 기반이 되는 보고서[8]에 따라, 효율계수 $\alpha = 0.2$, 동원률 $\beta = 1/3$ 으로 설정한다. 주간 시작시간을 8시부터 시작해서 주간작업시간을 10시간을 진행한다. 주간 시작시간을 T_{w_0} , 주간작업시간을 T_w , 사건이 시작된 시각을 T_0 로 표기한다. 작업시간을 8시간으로 산정하므로, 1시간마다 $\gamma = 0.8$ 의 효율을 보인다고 가정하여 식 5를 도출한다.

$$f(\mathbf{x}, i) = \min \left\{ t \in \mathbb{N} \mid \frac{Q_i}{3} - \alpha\beta\gamma \sum_{j=1}^n \max(t - g(t_{ij}, T_0 + t_{ii}), 0) x_j \leq 0 \right\} \quad (5)$$

3. 유전알고리즘을 이용한 방제자원배치 최적화

본 연구에서 방제자원배치 최적화를 위해 설계한 유전 알고리즘의 동작은 다음과 같다[9]. 자원배치 문제를 해결하는 유전알고리즘이 실행되면, 한 세대(population)에서 두 개의 서로 다른 배치를 선택(selection)하고, 두 개의 배치에서 교차(crossover) 연산을 통해 새로운 배치를 만든다. 또한, 임의의 확률로 할당량을 변화 시키는 변이(mutation)과정을 지나 새롭게 만들어진 배치는 주어진 제약조건을 만족하도록 수선(repair)하는 과정을 거쳐 문제 풀이에 대하여 적합도(fitness)를 평가(evaluation)받고, 이전 세대의 가장 좋은 배치 방법과 비교하여 적합도가 더 높으면 대체(replacement)된다[10]. 고안된 자원배치 최적화를 위한 유전알고리즘은 Fig. 1처럼 나타낼 수 있으며, 고안된 유전알고리즘은 Table 2와 같이 파라미터를 갖는다.

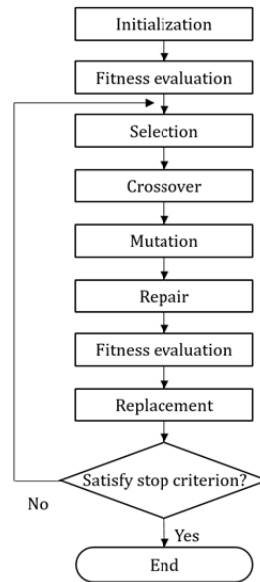


Fig. 1. Flowchart of our genetic algorithm for resource allocation optimization

Table 2. Parameters information of our genetic algorithm

| Parameter | Value |
|------------------------|-------|
| Max. generation number | 6,000 |
| Population size | 100 |
| Crossover rate | 0.7 |
| Mutation rate | 0.002 |

4. k-평균 클러스터링 분석

k-평균 알고리즘[11]은 군집분석[12]을 위해 대표적으로 사용되는 기법으로, 각 군집의 중심(μ)까지 유클리드 거리 차이의 분산(V)을 최소화하며, 데이터를 k 개의 군집(S)으로 나눈다. 해당 알고리즘을 수식으로 나타내면 식 6과 같다.

군집 분석을 위해 장비 동원 속도를 10 knot로 고정

한 후, 오염사고 발생 시각에 따라 유전알고리즘으로 최적화된 데이터를 10,000 개씩 수집하였다. 우리는 별도의 실험을 통해 k 가 2일 때, 클러스터가 가장 뚜렷하게 보임을 확인했다. 예상 최대유출량이 가장 많은 지역인 대신, 여수, 울산을 기준으로 할당된 배치량의 범주에 따라 군집화하면 Fig. 2와 같이 나누어진다.

$$V = \sum_{i=1}^k \sum_{j \in S_i} |x_j - \mu_i|^2 \quad (6)$$

5. 새몬맵핑을 이용한 분포도 분석

나누어진 군집들을 이용한 분포도 분석을 위해, 각 군집에서 1,000 개의 데이터를 샘플링하여 새몬맵핑[13]을 적용하였다. 새몬맵핑은 고차원 데이터의 인스턴스간 거

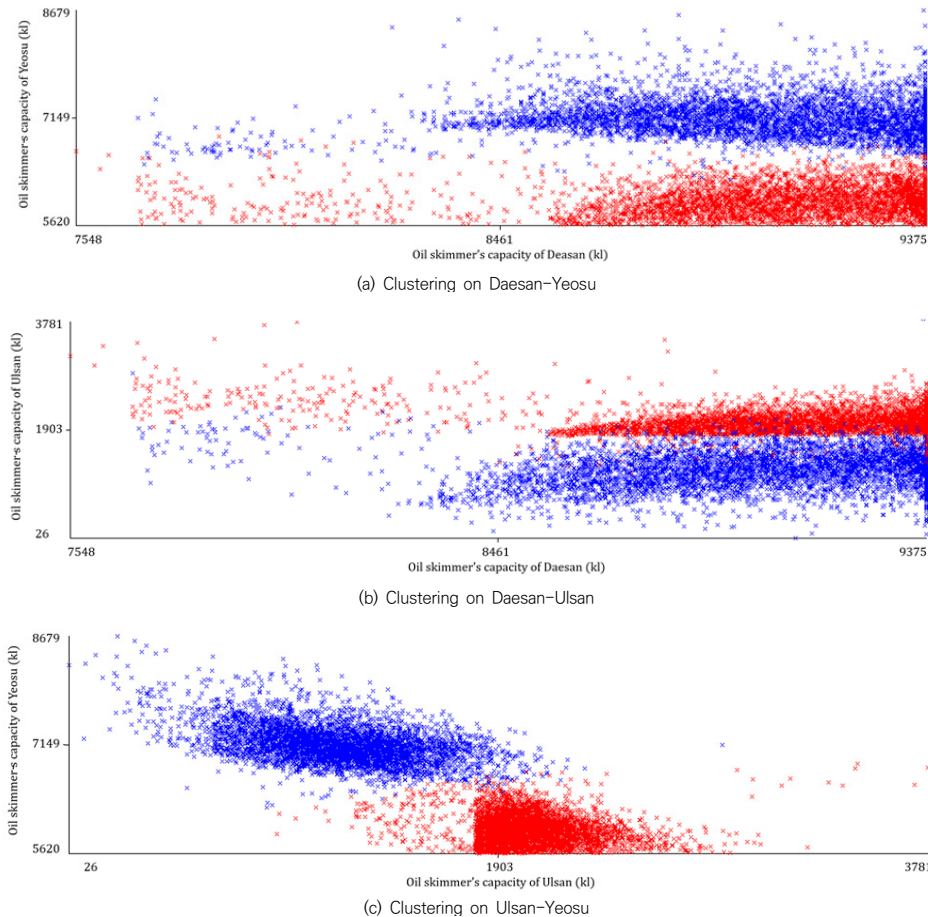


Fig. 2. Clustering results when oil spill accident

리를 바탕으로 저차원으로 사영하는 기법이며, 식 7처럼 나타낼 수 있다. δ_{ij} 가 원래 공간에서의 i 번째와 j 번째 인스턴스 사이의 거리이고, d_{ij} 가 사영된 거리일 때, 식 7을 최소화하는 것을 목표로 한다.

$$E = \frac{\sum_{i=1}^{n-1} \sum_{j=i+1}^n (\delta_{ij} - d_{ij})^2 / \delta_{ij}}{\sum_{i=1}^{n-1} \sum_{j=i+1}^n \delta_{ij}} \quad (7)$$

우리는 유클리드 거리를 기반으로 새몬맵핑을 적용하여 16 차원의 배치안 데이터를 2 차원으로 사영하였다. 또한 밀집도를 기반으로 Fig. 3과 같은 2 개의 군집을 확인하였으며, 군집의 평균 배치량과 군집 외의 주변 값인 A-D의 배치안을 분석한 결과는 Table 3과 같다. 배치안을 시물레이션하여 얻은 평균 작업시간을 평가했을 때, 작업에 소요되는 시간이 적을수록 우수한 배치안이라 볼 수 있다. 시물레이션 결과, 군집에 속하지 않는 주변값들보다 군집에 속하는 배치안들의 시물레이션 결과가 더 좋은 것을 확인하였다.

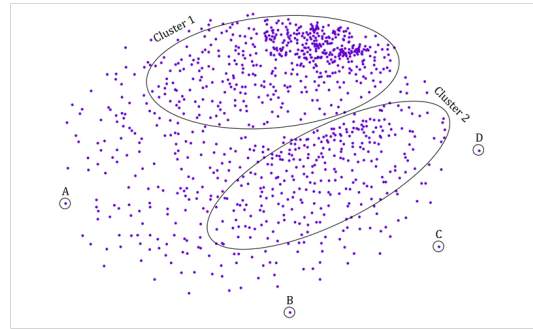


Fig. 3. Projection results by Sammon's mapping

6. 결론

우리는 메타 휴리스틱의 대표적인 기법인 유전알고리즘을 사용하여 자원 배치를 최적화하는 방법을 제안하고, 10,000 개의 최적화된 배치안을 수집하여 k -평균 알고리즘과 새몬맵핑을 이용하여 분포도를 분석하였다. 최적화된 배치안을 다양화 하고 분포도를 분석한 결과, 최대 유출지역인 대산, 여수, 울산의 배치량에 따라 2 가지의 패턴으로 나누어진다는 것을 알 수 있었다. 또한, 해당 패턴들에 속하는 배치안들이 좋은 시물레이션 결과를 보였다. 본 연구에서 진행한 최적화된 배치안을 다양화하여 분석한 결과들을 바탕으로, 시물레이션을 대체할 수 있는 근사모델[14, 15, 16]을 통해 속도적 이점이 있는 유전알고리즘을 설계할 수 있을 것이다.

Table 3. Distribution analysis of optimal resource allocations

| | Cluster1 | Cluster 2 | A | B | C | D |
|---------------------------|---------------|---------------|---------------|---------------|---------------|---------------|
| Incheon | 1701.8 | 1495.4 | 1643 | 1772 | 1743 | 1689 |
| Pyeongtaek | 194.2 | 228.6 | 248 | 250 | 227 | 250 |
| Deasan | 8636.8 | 9241.4 | 8789 | 9206 | 9183 | 8918 |
| Gunsan | 771.4 | 781.0 | 792 | 792 | 792 | 792 |
| Mokpo | 1641.4 | 1232.0 | 1564 | 926 | 1183 | 1202 |
| Wando | 76.2 | 109.8 | 125 | 118 | 98 | 125 |
| Yeosu | 5981.4 | 5851.0 | 5646 | 6004 | 5917 | 6048 |
| Jeju | 126.0 | 155.6 | 167 | 98 | 119 | 167 |
| Seogwipo | 59.6 | 104.2 | 64 | 105 | 75 | 44 |
| Tongyeong | 334.8 | 353.8 | 355 | 355 | 328 | 293 |
| Changwon | 225.8 | 245.0 | 250 | 250 | 217 | 250 |
| Busan | 503.6 | 457.0 | 522 | 470 | 522 | 522 |
| Ulsan | 2066.8 | 2004.8 | 2052 | 1926 | 1865 | 2020 |
| Pohang | 114.2 | 140.0 | 167 | 167 | 141 | 64 |
| Donghae | 62.0 | 90.8 | 105 | 50 | 79 | 105 |
| Sokcho | 4.0 | 9.6 | 11 | 11 | 11 | 11 |
| Average work hours | 17.340 | 17.345 | 17.340 | 17.368 | 17.317 | 17.385 |

REFERENCES

- [1] G. Tyagi, R. Singh & A. Hussain. (2019, Apr). Applications of genetic algorithm in water resources management and optimization. *International Conference on Advanced Computing and Software Engineering*, (pp. 137-143).
- [2] L. Wang, S. Guo, X. Li, B. Du & W. Xu. (2018). Distributed manufacturing resource selection strategy in cloud manufacturing. *The International Journal of Advanced Manufacturing Technology*, 94(9-12), 3375-3388. DOI: 10.1007/s00170-016-9866-8
- [3] J. Pearl. (1984). *Heuristics: Intelligent Search Strategies for Computer Problem Solving*. United States : Addison-Wesley Longman Publishing Co.
- [4] C. Blum & A. Roli. (2003). Metaheuristics in combinatorial optimization: overview and conceptual comparison. *Association for Computing Machinery Computing Surveys*, 35(3), 268-308. DOI : 10.1145/937503.937505
- [5] Z. S. Dai & X. L. Wang. (2006). Optimal resource allocation on grid systems for maximizing service reliability using a genetic algorithm. *Reliability Engineering & System Safety*, 91(9), 1071-1082. DOI : 10.1016/j.ress.2005.11.0008
- [6] V. Lavric, P. Iancu & V. Pleşu. (2004). Optimal water system topology through genetic algorithm under multiple contaminated-water sources constraint. *Computer Aided Chemical Engineering*, 18, 433-438. DOI: 10.1016/S1570-7946(04)80138-X
- [7] C. Guerrero, I. Lera & C. Juiz. (2018). Genetic Algorithm for Multi-Objective Optimization of Container Allocation in Cloud Architecture. *Journal of Grid Computing*, 16(1), 113-135. DOI : 10.1007/S10723-017-9419-X
- [8] J. H. Yun et al. (2009). *A Study on Practical Strategies for Estimating the National Control Ability of Oil Spill Control*. South Korea: Korea Coast Guard.
- [9] H. J. Kim, J. Lee, J. H. Yun & Y. H. Kim. (2019, July). Optimal equipment assignment for oil spill response using a genetic algorithm. *Genetic and Evolutionary Computation Conference Companion*. (375-376).
- [10] H. J. Kim & Y. H. Kim. (2019). Optimization of weapon target assignment using a genetic algorithm. *Journal of Korean Institute of Intelligent Systems*, 29(3), 176-181. DOI : 10.5391/JKIS.2019.29.3.176
- [11] K. W. C. Cardie & S. R. S. Schroedl. (2001). Constrained *k*-means clustering with background knowledge. *International Conference on Machine Learning*. (577-584).
- [12] H. J. Kim & Y. H. Kim. (2019, Dec). Distribution analysis of optimal resource allocation for oil spill response. *Korean Institute of Intelligent Systems Conference*. (193-194).
- [13] J. W. Sammon. (1969). A nonlinear mapping for data structure analysis. *Institute of Electrical and Electronics Engineers Transactions on Computers*, 18(5), 401-209. DOI : 10.1109/T-C.1969.222678
- [14] Y. Jin. (2011). Surrogate-assisted evolutionary computation: recent advances and future challenges. *Swarm and Evolutionary Computation*, 1(2), 61-70. DOI : 10.1016/j.swevo.2011.05.001
- [15] D. P. Yu & Y. H. Kim. (2019, July). Predictability on performance of surrogate-assisted evolutionary algorithm according to problem dimension. *Genetic and Evolutionary Computation Conference Companion*. (91-92).
- [16] H. J. Kim & Y. H. Kim (2020), *Optimal resource allocation using a genetic algorithm: application to weapon-target assignment and equipment assignment for oil spill response*. Doctoral dissertation. Kwangwoon University, Seoul.

김혜진(Hye-Jin Kim)

[정회원]



- 2018년 2월 : 광운대학교 컴퓨터소프트웨어학과(공학사)
- 2020년 2월 : 광운대학교 컴퓨터과학과(공학석사)
- 관심분야 : 인공지능, 유전알고리즘, 기계학습
- E-Mail : ova22@gmail.com

김용혁(Yong-Hyuk Kim)

[정회원]



- 1999년 2월 : 서울대학교 전산과학과(이학사)
- 2001년 2월 : 서울대학교 전기컴퓨터공학부(공학석사)
- 2005년 2월 : 서울대학교 전기컴퓨터공학부(공학박사)
- 2005년 3월 ~ 2007년 2월 : 서울대학교 반도체공동연구소 연구원
- 2007년 3월 ~ 2017년 2월 : 광운대학교 컴퓨터소프트웨어학과 조교수/부교수
- 2017년 3월 ~ 현재 : 광운대학교 소프트웨어학부 교수
- 관심분야 : 최적화, 진화연산, 지식공학
- E-Mail : yhdgly@kw.ac.kr