An Adaptive Estimation of Distribution Algorithm for Multipolicy Insurance Investment Planning

Wen Shi, *Student Member, IEEE*, Wei-Neng Chen🆔, *Senior Member, IEEE*, Ying Lin, *Member, IEEE*, Tianlong Gu, Sam Kwong🆔, *Fellow, IEEE*, and Jun Zhang🆔, *Fellow, IEEE*

Abstract—**Insurance has been increasingly realized as an important way of investment and risk aversion. Fruitful insurance products are launched by insurers, but there is little research on how to make a proper insurance investment plan for a specific policyholder given different kinds of policies. In this paper, we aim to propose a practical approach to multipolicy insurance investment planning with a data-driven model and an estimation of distribution algorithm (EDA). First, by making use of the insurance data accumulated in the modern financial market, an optimization model about how to choose endowment and hospitalization policies is built to maximize the yearly profit of insurance investment. With the model parameters set according to the real data from insurance market, the resulting plan is practical and individualized. Second, as the optimal solution cannot be achieved by mathematical deduction under this data-driven model, an EDA is introduced. To adapt the EDA for the considered problem, the proposed EDA is mixed with both the continuous and discrete probability distribution models to handle different kinds of variables. In addition, an adaptive scheme for choosing suitable distribution models and an efficient constraint handling strategy are proposed. Experiments under different conditions confirm the effectiveness and efficiency of the proposed model and method.**

Index Terms—**Data-driven, endowment insurance, estimation of distribution algorithm (EDA), hospitalization insurances, mixed-variable optimization.**

I. INTRODUCTION

INSURANCE is an important way for people to cope with risk. With the development of modern economy, more and more people realize the significance of insurance and consider it a key part of their investment plan [1], [2]. As the insurance industry grows rapidly in recent years, insurance products and behaviors become more diverse and sophisticated. It becomes increasingly important and challenging to study the actuarial models of insurance.

In the literature, various mathematical models for insurance have been developed, most of which focused on insurance pricing on behalf of the insurers, e.g., insurance companies. In the 1970s, Brennan and Schwartz [3] proposed the first pricing model that considered the equity-linked life insurance policies with an asset value guaranteed. Since then, researchers have developed a plethora of models to address the problem of insurance pricing under different circumstances. Some of the representative works include the pricing of an endowment insurance policy paid by annual premium [4], the models for equity-linked life insurance policies [5]–[7] and fixed income linked life insurance policies [8]. Also, there are dozens of researchers studying the pricing model of nonlife policies. Please refer to [9] for a detailed survey on such pricing models.

In contrast to the pricing models for insurers, the research into insurance investment models on behalf of the policyholders is far less matured. In the literature, the investments of insurance were mostly considered as a complementary part of the entire property investment plan. The first insurance investment model was proposed by Hakansson [1], which considered life insurance as a part of the property investment plan and established a discrete model to calculate the return of life insurance. Later the model was extended to a continuous form for enhancing its practical value [2], as the investment process into insurance is actually a continuous procedure. Based on these two models, a number of improved models have been developed. Pliska and Ye [10] studied the optimal investment strategy for a wage earner with uncertain lifetime. The model was then extended to take risky assets [11]–[13] and different kinds of life insurance products [14] into account. Besides the above models that consider the investment strategy of a single person, investment models for the collective situation [15], [16] were also developed. However, all the above models share a vital defect. That is, only one type of insurance policies, i.e., life insurance, is considered. In fact, to fulfill desires of different policyholders, the modern insurance market offers many different types of insurance policies, including but not limited to endowment policies, hospitalization policies, property policies, auto policies, etc. To make the insurance investment models better suit the modern insurance

W. Shi, W.-N. Chen, and J. Zhang are with the School of Computer Science and Engineering, South China University of Technology, Guangzhou 510006, China, and also with the Guangdong Provincial Key Laboratory, Computational Intelligence and Cyberspace Information, Guangzhou 510006, China (e-mail: cwnraul634@aliyun.com; junzhang@ieee.org).

Y. Lin is with the Department of Psychology, Sun Yat-sen University, Guangzhou 510006, China.

T. Gu is with the School of Computer Science and Engineering, Guilin University of Electronic Technology, Guilin 541004, China.

S. Kwong is with the Department of Computer Science, City University of Hong Kong, Hong Kong.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

market and better satisfy the practical need of policyholders, it is necessary to take more types of policies into account.

In order to overcome the above deficiency, this paper intends to establish an insurance investment model that considers different kinds of endowment and hospitalization policies. However, the fact that different insurance policies calculate their premiums and profits in different ways increases the complexity of the proposed model and brings up the following two challenges.

On the one hand, in the case that only life insurance is considered, it is possible to use mathematical expressions to deduce the price and the profit of an insurance policy [10]–[14]. As for the case of multipolicy insurance investment, since many factors (e.g., wealth condition, health condition, age, etc.) must be taken into account, it is difficult or even impossible to deduce the price and the profit of each policy mathematically. For instance, to mathematically model a multipolicy insurance investment plan that considers the health condition of a policyholder, a mathematical expression of the health condition is needed at first. However, since it is hard to establish an accurate mathematical expression to assess the health condition of the policyholder, the establishment of the mathematical model for the multipolicy insurance plan becomes difficult. With the rapid development of insurance business, a wealth of insurance data (settings of insurance policies, demographic information of policyholders, real insurance cases, etc.) has been accumulated. Instead of using mathematical expressions, it might be more practical and promising to build the insurance investment model based on data. Inspired by this idea, this paper proposes a data-driven multipolicy insurance investment model.

On the other hand, the simplicity of the existing investment models for life insurance makes it possible to derive the optimal investment plan by mathematical deduction. However, the proposed multipolicy insurance investment model becomes so complex that it may not have the mathematical properties (e.g., continuity or differentiability) required by traditional optimization methods based on mathematical deduction. Therefore, this paper proposes the use of estimation of distribution algorithms (EDAs) [17], a simple yet effective evolutionary algorithm [18]–[22], to address the proposed multipolicy insurance investment model. There are two advantages of utilizing EDA. First, EDA is a stochastic optimization algorithm based on statistical learning theory, and thus has inherent coherence with the proposed insurance investment model established based on statistical data. Second, it has been shown that EDA has good global search capability, which makes it promising for optimization of the investment plan involving multiple insurance policies.

In order to tailor the original EDA to the proposed insurance investment model, the following improvements are made. First, since the multipolicy insurance investment model has both of continuous and discrete decision variables, a mixed variable framework that enables EDA to handle different types of variables simultaneously is proposed in this paper. Second, to avoid the premature convergence, an adaptation strategy that can choose the probabilistic model in EDA according to the performance is developed. Third, to ensure that the

insurance investment plan obtained is feasible, a novel constraint handling strategy is designed and embedded. Finally, a local search strategy is utilized to further improve the search efficiency. Experiments under different conditions confirm the effectiveness and efficiency of the proposed model and method.

The rest of this paper is organized as follows. Section II briefly introduces the background of insurance investment models and EDAs. Section III formally defines the proposed multipolicy insurance investment model. Section IV introduces the details of the proposed EDA for optimization of the proposed model. Section V presents a series of experiments to prove the effectiveness and efficiency of the proposed model and method. Finally, the conclusion of this paper is drawn in Section VI.

II. BACKGROUND

A. Insurance Investment Model

Insurance is a kind of investment strategy that can help the policyholder to avoid risks. In particular, life insurance takes the life expectancy of the insured as the subject of insurance, and the payment condition is the survival or death of the insured. Most existing works about insurance investment only considered a single life insurance policy [10]–[13], in which a continuous investment model is employed and the return of the life insurance, $Z(t)$, is calculated according to the following formula:

$$Z(t) = \frac{I(t)}{\eta(t)} \tag{1}$$

where $I(t)$ is the amount that the policyholder purchases at time t and $\eta(t)$ is the premium-payout ratio at time t.

With the rapid development of insurance business, a variety of life insurance policies with different characteristics have been developed. The necessity of studying the investment strategy involving more than one life insurance policy thus becomes more and more obvious. Mousa $et\ al.$ [14] proposed the first and, so far, the only one insurance investment model that considered different kinds of life insurance policies. Yet the difference in different policies was limited to their premium-payout ratios. The model considering k life insurance policies was thus extended to

$$Z(t) = Z_1(t) + \cdots + Z_k(t) = \frac{I_1(t)}{\eta_1(t)} + \cdots + \frac{I_k(t)}{\eta_k(t)}. \tag{2}$$

Here, $Z_i(t)$, $I_i(t)$, and $\eta_i(t)$, $i = 1, 2, \ldots k$, denote the return of the ith policy, the amount of purchasement of the ith policy at time t, and the premium-payout ratio of the ith policy at time t, respectively. With the model, it is obvious that the policy with the lowest premium-payout ratio at time t would be chosen, while the purchase amount of the other policies would be zero.

From the above, it can be known that the previous models have many limitations on the type and the amount of insurance policies in consideration. In fact, nowadays, few companies offer life insurance policies that only have death compensation. Most companies provide endowment policies that add
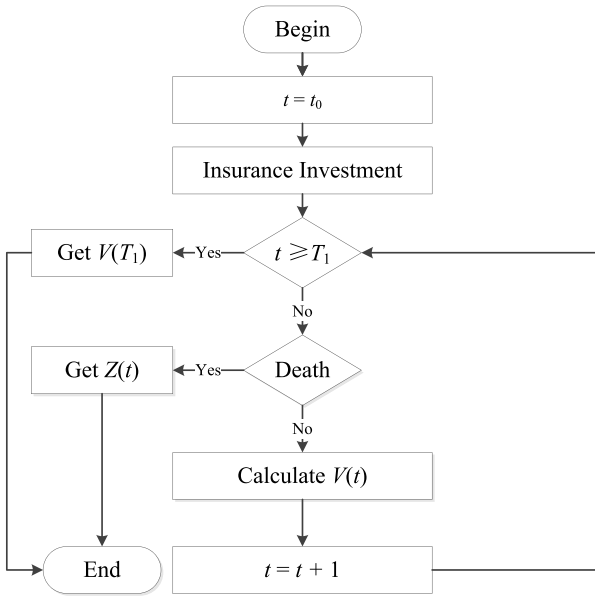
Fig. 1. Flowchart of insurance investment model for endowment policies.

Fig. 2. Flowchart of EDAs.

the function of saving to life insurance [23]. As the calculation of the return of endowment policies is more complex, the previous models are no longer suitable for studying. To show the insurance investment model for endowment policies, the flowchart is given in Fig. 1.

Besides the endowment policies, this paper also takes a kind of popular nonlife insurances, the hospitalization policies [24], into account. A hospitalization policy is an insurance contract that helps the policyholder to avoid the risk of incurring medical expenses when getting certain kind of diseases. Suppose that the premium of a hospitalization policy is $z(t)$, if the policyholder gets sick with certain disease specified in the policy at time t, the medical expenses $Y(t)$ will be reimbursed by the insurer, that is, the policyholder can get a return of $Y(t)$. If the policyholder remains healthy, he will get no return. In summary, the total return $R(t)$ can be calculated as

$$R(t) = \begin{cases} -z(t), & \text{if the policyhoder is fine} \\ Y(t) - z(t), & \text{otherwise.} \end{cases} \quad (3)$$

Taking multiple endowment and hospitalization policies into account significantly improves the practicality of the insurance investment model. However, it also makes the model more complex, which increases the difficulty of finding the optimal investment plan of the model. Considering that the model may not have the mathematical properties required by traditional optimization methods, usage of a new kind of optimizer, the EDA, is proposed.

B. Estimation of Distribution Algorithm

EDA [17] is an evolutionary algorithm based on statistical learning theory. As shown by the flowchart in Fig. 2, EDA starts from initializing a population, in which each individual represents a candidate solution to the problem. Based on the existing population, it utilizes the statistical learning method to estimate a probability model for describing the distribution of promising solutions in the search space. Then the
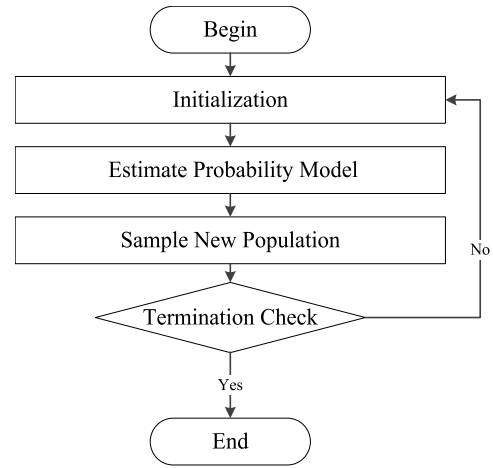
probability model is sampled to generate a new population. By repeating the steps of estimating and sampling the probability model of promising solutions, EDA can evolve the population toward the optimum of the problem.

Building a probability model that can accurately describe the distribution of promising solutions in the search space is the key to enhance the efficiency of EDAs. Considering the characteristics of different kinds of problems, EDAs with different model estimation methods are developed. Population-based incremental learning (PBIL) [25], univariate marginal distribution algorithm (UMDA) [26], and compact genetic algorithm (cGA) [27] are proposed to deal with the optimization of variable independent binary coding problems. Mutual information maximization for input clustering [28], combining optimizers with mutual information trees [29], and bivariate marginal distribution algorithm [30] are proposed to solve the optimization problems of bivariate correlation binary coding. Factorized distribution algorithms [31], extended cGA [32], and Bayesian optimization algorithm [33] are proposed to deal with the optimization of multivariate correlation binary coding problems. PBILc [34], UMDAc [35], and stochastic hill climbing with learning by vectors of normal distributions [36] are proposed for variable independent real-number coding problem. Estimation of multivariate normal algorithm [37] and IDEA [38] are proposed to solve the multivariate correlation real-number coding problem. In recent years, many EDA variants have been proposed [39]–[44] and it is extensively applied to solve many problems, such as multimodal problems [45], [46] and multiobjective problems [47].

However, none of the above EDAs are sufficiently suitable for the insurance investment problem considered in this paper. The reasons are as follows. First, the multipolicy insurance investment model proposed in this paper consists of both discrete and continuous variables. To ensure precision and efficiency of optimization, instead of encoding all of the variables into a binary vector or a real-number vector, it is more desired to adopt a mixed coding scheme. Yet all the above EDAs are designed using a single (either binary or real-number) coding scheme. Second, there are specific constraints among the

variables. The problem cannot be easily transformed into any multivariate correlation problems dealt by the above EDAs. In order to conquer the above problems, this paper develops an adaptive EDA (AEDA) to address the proposed multipolicy insurance investment model, which will be elaborated in Section IV.

III. MULTIPOLICY INSURANCE INVESTMENT MODEL

This paper focuses on optimization of insurance investment plan that involves endowment and hospitalization policies. For simplicity, we only consider the situation that the policyholder is also the insured.

In the proposed model, different kinds of endowment and hospitalization policies are considered. One can purchase as many kinds of endowment policies and invest as much as he wants, but the investment for each payment period must fulfill the minimum request. As for the hospitalization policies, the premium is determined based on the coverage scope as well as the holder's situation. Only one hospitalization policy can be purchased, and to keep the policy effective, the policyholder must pay the same amount of premium every year. In exchange, he no longer needs to pay the medical expenses incurred by the diseases covered by the policy.

Given the initial wealth and the policyholder's initial age, the optimization objective is to maximize the statistical value of the yearly profit by tuning the amount of investment for each payment period of the endowment policies and the choice of hospitalization policy.

A. Parameters of the Model and Their Settings

From the above description, it can be seen that a number of parameters must be defined at first to make the proposed model work. The parameters are listed as follows.

1) Parameters Related to Endowment Policies: The following parameters are needed to compute the profit of the ith endowment policy with the jth payment period ($i = 1, 2, \ldots, n_1, j = 1, 2, \ldots, m_i$):

1) $\eta_{1ij}(t_0, t)$ return rate at time t if invested at t_0;
2) $\eta_{2ij}(t_0, t)$ death compensation rate at time t if invested at t_0;
3) q_{ij} latest time to purchase;
4) l_{ijt} minimum amount of investment;
5) T_1 guarantee period.

2) Parameters Related to Hospitalization Policies: The following parameters are needed to analyze the investment efficiency of the hospitalization policies:

1) $\delta_k(t_0)$ premium of policy k at time t_0, $k = 1, 2, \ldots, n_2$;
2) A_k coverage scope of policy k, $k = 1, 2, \ldots, n_2$;
3) z_s medical expense for diseases with the sth serious degree;
4) T_2 latest time to buy a hospitalization policy.

3) Parameters Related to the Policyholder: The following parameters are needed to design an individualized investment plan:

1) I_0 initial wealth;
2) $I(t)$ annual income at time t;
3) t_0 initial age;

4) $p(t)$ mortality rate at time t;
5) $p_s(t)$ incident rate for diseases with the sth serious degree at time t.

For a policyholder who wants to utilize the proposed model to derive the optimal investment plan, the parameters related to endowment policies and hospitalization policies should be given in advance. Some of the parameters related to the policyholder, i.e., $I_0, I(t)$, t_0, should also be given to obtain an individualized insurance investment plan. The remaining parameters, i.e., $p(t)$ and $p_s(t)$, should be set based on the statistical data as below.

Let $T(x)$ be the number of years a person at the age of x can live. The probability for the person to survive after τ years, which is denoted as $_\tau p_x$, is calculated as

$$_\tau p_x = \Pr(T(x) > \tau) = \Pr(X > x + \tau | X > \tau) = \frac{s(x + \tau)}{s(x)} \tag{4}$$

where $\Pr(A)$ indicates the probability that A will occur, while $\Pr(A|B)$ is the probability that A will occur given that event B has occurred. $s(x + \tau)$ is the number of survivors at time $(x + \tau)$ and $s(x)$ is the total number of survivors at x. Based on the above, the mortality rate at time $(x + \tau)$ for a person at the age of x can be computed as

$$p(x + \tau) =_{\tau+1} p_x -_\tau p_x = \frac{s(x + \tau + 1) - s(x + \tau)}{s(x)}. \tag{5}$$

To complete the above derivation, $s(\cdot)$ must be known in advance. Therefore, it is necessary to estimate the age distribution of mortality over the people whose health conditions and health habits are similar with the policyholder. Such estimation can be achieved by collecting and analyzing historical data.

The incidence rates for diseases of the sth serious degree are derived as

$$p_s(x + \tau) = \frac{i_s(x + \tau)}{s(x + \tau)} \tag{6}$$

where $i_s(x + \tau)$ is the number of survivors who have diseases of degree s at $x + \tau$. To conduct the above derivation, $i_s(\cdot)$ and $s(\cdot)$ must be known in advance. It is thus necessary to estimate the incident age distribution of patients whose health conditions and health habits are similar to the policyholder. Such estimation can also be achieved based on historical data.

In the literature, only death compensation rate and mortality rate are considered as parameters of the insurance investment models [10]–[16]. The difference in insurance policies and the policyholder's specific situations is largely neglected. In this paper, we propose to consider all of the parameters mentioned above. By doing so, the insurance investment model becomes more realistic, and the insurance investment plan can be optimized in an individualized way.

B. Premises and Assumptions

To facilitate the establishment of the optimization model, the following premises and assumptions are made according to

existing literatures of insurance investment and some insurance policies in real life [10]–[14], [48].

1) Suppose that there are n_2 hospitalization policies. All of the diseases in concern are classified into n_2 groups regarding their serious degrees, and the groups are indexed in descending seriousness. Assume that the coverage scope of the ith hospitalization policy includes all the diseases in the first i group(s) ($i = 1, 2, \ldots, n_2$). In other words, the first policy only covers the most serious diseases, whilst the n_2th policy covers all the diseases in concern. The more diseases a policy covers, the higher its premium is. Note that the premium of a hospitalization policy is determined at the time of purchase. The policyholder must pay the same amount of premium every year till the payment period is completed.

2) The cash values of endowment policies cannot be cashed in to purchase new endowment policies or to pay the medical expense for the diseases outside the coverage of the selected hospitalization policy. However, the cash values can be used to pay the premium of the hospitalization policy.

3) Assume that the policyholder does not need to withdraw any money from his investment into insurance. Surrender policies are thus not considered in the proposed model. Each insurance policy purchased by the policyholder will stay effective until its guarantee period is completed.

4) Assume that the policyholder's death always occurs at the beginning of a year. Therefore, the cash value of the endowment policies and the possible medical expenses in the year of the policyholder's death are not considered when computing the yearly profit.

5) Assume that the medical expenses for the diseases with the same degree of seriousness are constant, irrespective of age or general health conditions. The more serious a disease is, the higher the medical expense will be.

6) The policyholder begins to profit from the purchased endowment policies one year afterwards.

C. Objective Function

Suppose that there are n_1 endowment policies and n_2 hospitalization policies in the market. Each endowment policy i has m_i payment periods ($i = 1, 2, \ldots, n_1$). According to the previous sections, the optimization problem of the insurance investment plan can be formulated as

$$\max J(t; X, \beta) \tag{7}$$

where $J(t; X, \beta)$ is the statistical value of the profit at time t given the investment plan X of endowment policies and the selected hospitalization policy β. More specifically, X is an irregular matrix, in which each element x_{ij} indicates the amount of investment for the jth payment period of the ith endowment policy. $J(t; X, \beta)$ is calculated as

$$J(t; X, \beta) = [1 - p(t)]V(t) + p(t)Z(t) + C(t) \tag{8}$$

where $p(t)$ is the mortality rate of the policyholder at time t, $V(t)$, $Z(t)$, and $C(t)$ denote the cash value, the death

compensation (if occur), and the cash on hand at time t, respectively.

The cash value at time t is the sum of the current cash values of all the endowment policies, i.e.,

$$V(t) = \sum_{i=1}^{n_1} \sum_{j=1}^{m_i} V_{ij}(t) \tag{9}$$

where $V_{ij}(t)$, as calculated by (10), is the cash value of the ith endowment policy with the jth payment period at time t

$$V_{ij}(t) = \begin{cases} x_{ij}\eta_{1ij}(t_0, t), & \text{if } t = t_0 + 1 \\ V_{ij}(t-1) + x_{ij}[\eta_{1ij}(t_0, t) - \eta_{1ij}(t_0, t-1)], & \text{otherwise.} \end{cases} \tag{10}$$

In (10), t_0 is the initial age of the policyholder and $\eta_{1ij}(t_0, t)$ is the return rate of the ith endowment policy with the jth payment period. Note that the cash values will be used to pay the premium of the selected hospitalization policy, as long as the policyholder's age has not passed the latest investment time T_2 (i.e., $t \le T_2$). In that case, the cash values $V_{ij}(t)$ are taken out in descending order to pay the premium $\delta_\beta(t)$ of the hospitalization policy β till $\delta_\beta(t)$ is paid off. If $\delta_\beta(t)$ cannot be fulfilled with all the cash values taken out [i.e., $\delta\beta(t) > V(t)$], the hospitalization policy is considered unaffordable. The investment strategy thus becomes infeasible. For each endowment policy whose cash value has been taken out, $V_{ij}(t)$ is set as zero. That is, the accumulation of the cash value restarts from time $(t + 1)$.

Likewise, the death compensation at time t is the sum of the death compensations of all the endowment policies. Let $\eta_{2ij}(t_0, t)$ be the death compensation rate of the ith endowment policy with the jth payment period. The death compensation at time t is computed as

$$Z(t) = \sum_{i=1}^{n_1} \sum_{j=1}^{m_i} x_{ij}\eta_{2ij}(t_0, t). \tag{11}$$

The cash value at time t is equal to the balance at time $(t-1)$ plus the policyholder's income $I(t)$, and minus the total medical expenses. Let $p_s(t)$ be the incidence rate for diseases in the sth group at time t and z_s be the medical expense for the diseases in the sth group. The cash value at time t can be computed as

$$C(t) = \begin{cases} I_0, & \text{if } t = t_0 \\ C(t-1) + I(t) - \sum_{i=1}^{n_1} \sum_{j=1}^{m_i} v_{ij}x_{ij} - \sum_{s=1}^{n_2-\beta} p_s(t)z_s, & \text{otherwise} \end{cases} \tag{12}$$

where $v_{ij} \in \{0, 1\}$ is an indicator of whether or not the policyholder has finished the payment for the ith endowment policy with the jth payment period.

D. Constraints of the Model

In reality, investing insurance has some constraints. Such constraints are also manifested in the proposed model.

1) There is a latest time to invest in the jth payment period of the ith endowment policy, denoted as q_{ij}. That is, for

a policyholder whose age is greater than q_{ij}, it is not available for him to invest in the jth payment period of the ith endowment policy.

2) There is a minimum amount of the investment to the jth payment period of the ith endowment insurance at time t, denoted as l_{ijt}. That is, the investment amount to the jth payment period of the ith endowment insurance at time t should not be less than l_{ijt}.

3) There is a guarantee period for all the endowment policies, denoted as T_1. That is, after T_1, the insurance coverage period of the endowment policies is over, the cash value and the death compensation would not be accumulated any more.

4) There is a latest time to buy hospitalization policies, denoted as T_2. That is, the policyholder cannot buy any hospitalization policies after age T_2 if he has not bought any hospitalization policies before. But for a policyholder who has bought a hospitalization policy before T_2, he can continue to buy it after T_2.

5) In no case can a policyholder choose an investment plan that the total amount of cash value is not able to cover the premium of the hospitalization policy at any time.

6) The policyholder's cash in hand cannot be negative at any time otherwise the policyholder will go bankrupt.

IV. ADAPTIVE EDA APPROACH

The definition in Section III shows that the multipolicy insurance investment model is a complex optimization problem with multiple constraints and mixed variables. It is unlikely that the optimal solution can be obtained through traditional optimization methods based on mathematic computation. This paper proposes an AEDA to approximate the optimal solution of the problem. To ensure that the solution obtained is feasible, a novel constraint handling strategy is embedded into AEDA. To deal with the continuous and discrete variables in the problem simultaneously, AEDA adopts a mixed-variable framework to treat different types of variables separately. Besides, AEDA employs an adaptation strategy that can choose probability models automatically regarding their performances to avoid premature. A local search strategy is also employed to further improve the search efficiency. Detailed design of AEDA is illustrated in the following sections.

A. Initialization With Constraint Handling

Each individual in the population of AEDA represents a candidate insurance investment plan, which, according to Section II, can be written as $S = <X, \beta>$. In order to promote the search efficiency, it is better that the search starts from feasible solutions with adequate variety. To achieve this goal, AEDA initializes each individual by assigning random values to the variables in S such that the resulting investment plan satisfies the following constraints.

1) The value of β is an integer in the range of $[0, n_2]$.

2) For each element x_{ij} in X, the value of x_{ij} is not smaller than l_{ijt_0}, i.e., the minimum amount of investment

requested by the ith endowment policy with the jth period at time t_0.

3) For the endowment policies whose latest purchasing time q_{ij} is earlier than t_0, x_{ij} must be zero.

4) The sum of the elements in X, that is, the total amount of investment to endowment policies, cannot exceed the cash that the policyholder has on hand at any time.

Among the above four constraints, the first two can be satisfied by setting the ranges from which β and x_{ij} are sampled accordingly. The third constraint can be satisfied by comparing t_0 with q_{ij} before assigning a nonzero value to x_{ij}. The fourth constraint, on the other hand, is more difficult to deal with, as the amount of the cash on hand is uncertain except at time t_0.

To deal with the fourth constraint, a novel constraint handling process is proposed. In this process, the investment amount to each endowment policy and each payment period is initialized in sequence. To ensure that the total amount of investment does not exceed the cash the policyholder has, once the investment amount to a payment period of an endowment policy is decided, a corresponding amount should be subtracted from the cash the policyholder has in the following years. To avoid the situation that all the cash is invested to the first several policies for all individuals, the order of the policies is randomly perturbed for each individual. The detailed description of this constraint handling strategy is given below.

First, one of the endowment policy a and its payment period b is selected randomly from $a \in [1, n_1]$, $b \in [1, m_a]$.

Second, the upper bound of x_{ab} is computed as

$$u_{ab}(t_0) = \min\left\{\hat{C}(t_0), \hat{C}(t_0 + 1)/2, \ldots, \hat{C}(t_0 + y_{ab})/y_{ab}\right\}$$
(13)

where y_{ab} is the length of the payment period and $\widehat{C}(t)$, initialized by (14), is the expectation of the cash on hand at time t

$$\hat{C}(t) = \begin{cases} I_0, & \text{if } t = t_0 \\ \hat{C}(t-1) + I(t) - E_s(z_s), & \text{otherwise.} \end{cases}$$
(14)

Herein, I_0 is the policyholder's initial wealth, $I(t)$ is his income at time t, and $E_s(z_s)$ is the expectation of the policyholder's medical expense at time t. By setting x_{ab} as a random value not larger than $u_{ab}(t_0)$, the investment should be affordable at any time during the payment period. Note that the cash for $t = t_0, t_0 + 1, \ldots, t_0 + y_{ab}$ must be updated accordingly after setting x_{ab}. That is, after the value of x_{ab} is determined, the cash in hand should be taken out to pay that premium at each time t. Therefore, $\widehat{C}(t_0)$ should minus x_{ab}, $\widehat{C}(t_0 + 1)$ should minus $2x_{ab}$, and so on till the last year of the payment period when the cash in hand $\widehat{C}(t_0 + y_{ab})$ should minus $y_{ab}x_{ab}$. The above steps are repeated until the cash in hand can afford no more endowment policies.

The above initialization procedure for each individual $S^{(k)}$ in the population is summarized in Algorithm 1, where $k = 1, 2, \ldots, NP$ and NP is the population size.

B. Estimation of Probability Models

In each generation of the proposed AEDA, the best NP_{best} individuals in the current population are used to estimate the

---

Algorithm 1: Initialization

1: **for** $k = 1$ **to** NP **do**
2:      $\beta^{(k)} = Uniform\_int\,(0, n_2)$
3:      **Initialize** $\widehat{C}^{(k)}(t)$
4:      $A = \{1, 2, \ldots n_1\}$
5:      **for** $i = 1$ **to** n_1 **do**
6:          randomly choose a from A
7:          $A = A \setminus \{a\}$
8:          $B = \{1, 2, \ldots m_a\}$
9:          **for** $j = 1$ **to** m_a **do**
10:             randomly choose b from B
11:             $B = B \setminus \{b\}$
12:             **if** $t_0 \leq q_{ab}$ and $u_{ab}(t_0) \geq l_{abt_0}$
13:                 $x_{ab}^{(k)} = Uniform\,(l_{abt_0}, u_{ab})$
14:             **else**
15:                 $x_{ab}^{(k)} = 0$
16:             **end if**
17:             **for** $t_1 = t_0$ **to** $t_0 + y_{ab} - 1$ **do**
18:                 $\widehat{C}^{(k)}(t_1) = \widehat{C}^{(k)}(t_1) - x_{ab}^{(k)} \times (t_1 - t_0 + 1)$
19:             **end for**
20:             **for** $t_1 = t_0 + y_{ab}$ **to** T_1 **do**
21:                 $\widehat{C}^{(k)}(t_1) = \widehat{C}^{(k)}(t_1) - x_{ab}^{(k)} \times y_{ab}$
22:             **end for**
23:         **end for**
24:     **end for**
25: **end for**

---

probability models from which new individuals are sampled. The probability models of the continuous variables x_{ij} and the discrete variable β are built separately. In detail, for the investment amount x_{ij} of the jth payment period of the ith endowment policy, the probability model is established using the mean μ_{ij} and the standard deviation σ_{ij} calculated as

$$\mu_{ij} = \frac{\sum_{k=1}^{NP_{\text{best}}} x_{ij}}{NP_{\text{best}}} \tag{15}$$

$$\sigma_{ij} = \sqrt{\frac{\sum_{k=1}^{NP_{\text{best}}} (x_{ij} - \mu_{ij})^2}{NP_{\text{best}}}}. \tag{16}$$

The probability model can be the Gaussian distribution or the Cauchy distribution. An adaptation strategy is proposed to select between the two models according to their performances in the previous generations, as will be illustrated in the next section.

As for the choice β of hospitalization policies, a histogram method is used. For each available value i of β ($i \in \{0, 1, \ldots, n_2\}$), let count_i be the number of individuals whose value of β equals i in the current population and count_best_i be the number of individuals whose value of β equals i among the best NP_{best} individuals. The probability β_{i_p} for a new individual to assign i to β is calculated as

$$\beta_{i\_p} = \left[\sum_{k=0}^{n_2} \frac{\text{count\_best}_k}{\text{count}_k} \right]^{-1} \frac{\text{count\_best}_i}{\text{count}_i}. \tag{17}$$

By doing so, the choices that induce larger fitness in the current population are more likely to be used in the next generation.

Using the above strategy, AEDA can handle the continuous variables and the discrete variable in the proposed insurance investment model simultaneously. Besides, in this algorithm, the estimation of the probability model is based on statistical data. Thus, it is in accordance with the proposed model which is established based on statistical data as well.

C. Adaptive Model Selection

The probabilistic function of Gaussian distribution is a bell-shaped curve with thin tails on both sides. As a result, if the Gaussian distribution is used to model the distribution of promising solutions in the search space, most of the new individuals sampled from the model will be closed to the mean, causing decrease in the population diversity and thus early convergence. In contrast, the Cauchy distribution has relatively fat tails. Using the Cauchy distribution can therefore relieve the problem of early convergence, but also stems a new problem: slow convergence speed. In order to achieve a better balance between exploitation and exploration, this paper proposes an adaptation strategy, which chooses between the Gaussian distribution and the Cauchy distribution according to their performances in previous search.

The performance of each distribution is determined by the proportion of the promising individuals that use a kind of distribution to the total of individuals that use this kind of distribution. The distribution with the better performance should be utilized more when sampling the new population. In the first generation, since no information about the performance of the distribution can be obtained, half of the population uses the Gaussian distribution to generate offspring while the other half uses the Cauchy distribution. Therefore, the probability of choosing the Gaussian distribution to sample new population g_p is initialized as 0.5 and updated as follows:

$$g\_p = \left[\sum_{k=0}^{1} \frac{\text{top\_}d_k}{\text{count\_}d_k} \right]^{-1} \frac{\text{top\_}d_0}{\text{count\_}d_0}. \tag{18}$$

In (18), $\text{top\_}d_0$ and $\text{top\_}d_1$ is the number of individuals sampled from Gaussian distribution and Cauchy distribution in the best NP_{best} individuals, respectively. $\text{count\_}d_0$ and $\text{count\_}d_1$ is the number of offspring sampled from Gaussian distribution and Cauchy distribution, respectively. To avoid the situation that the probability of choosing one distribution becomes too large or too small, the value of g_p is confined to [0.1, 0.9].

D. New Population Generation

After the probability models are determined as in Sections IV-B and IV-C, a new population is generated by sampling these models. In this process, the individuals in new population must also satisfy the constraints listed in Section IV-A. Therefore, a constraint handling strategy similar to the initialization process is applied to ensure that the new individual is feasible. The detail process of generating a new individual is given below.

Step 1: Determine the value of β according to the histogram model built by (17). Initialize the cash value in hand according to (14).

Step 2: Randomly select an endowment policy a and its payment period b. Sample the corresponding model $N(\mu_{ab}, \sigma_{ab})$ (if the Gaussian distribution is selected) or $C(\mu_{ab}, \sigma_{ab})$ (if the Cauchy distribution is selected) to generate the investment x_{ab} for the new individual.

Step 3: Examine whether x_{ab} fulfills the related constraints. That is, x_{ab} cannot be less than the minimum request l_{abt_0} or exceed the cash in hand $u_{ab}(t_0)$. If x_{ab} exceeds $u_{ab}(t_0)$, it is set as $u_{ab}(t_0)$. If x_{ab} is in the range of $[l_{abt_0}/2, l_{abt_0})$, it is set as l_{abt_0}. If x_{ab} is less than $l_{abt_0}/2$, it is set as 0.

Step 4: Update the amounts of the cash in hand in the following years accordingly. More specifically, the maximum amount of investment for the other endowment policies should be recalculated using (13).

Step 5: Repeat steps 2–4 until the investment plan X for the endowment policies is fully determined. Then in combination with the choice of hospitalization policy determined in step 1, a feasible new individual is generated.

Repeat the above process till NP new solutions are generated. Algorithm 2 gives the pseudocode of generating a new population.

E. Local Search Strategy

To further improve the efficiency of the proposed AEDA, a local search strategy is employed. In this process, several individuals are generated by sampling the neighborhood around the best-so-far individual. The sampling operation is similar to the one developed in Section IV-D, which samples a new individual from the probability model built in advance.

For each individual generated in the local search process, the value of β is decided according to the histogram model built in Section IV-C. As for the value of $X = (x_{ij})$, $(i = 1, 2, \ldots, n_1, j = 1, 2, \ldots, m_i)$, it is generated by sampling the following distribution:

$$X \sim N\left(X^{\text{best}}, \sigma_l\right) \qquad (19)$$

where X^{best} denotes the best-so-far individual and is updated per generation. $\sigma_l \in (0, x_{ij}^{\text{best}})$ is the scale factor of this distribution and set as 10 according to the magnitude of the solution in this paper. $N(\mu, \sigma)$ is a normal distribution with expectation μ and standard deviation σ.

F. Overall Procedure

Based on the above operators, the overall procedure of AEDA can be described as follows.

Step 1 (Initialization): This process generates an initial population, in which each individual represents a feasible insurance investment plan.

Step 2: (Estimation of the Probability Models): This process estimates the probabilistic models of all the decision variables by learning from the best NP_{best} individuals in the current population.

Step 3 (Adaptive Model Selection): For each individual in the population, this process automatically decides whether the

---

Algorithm 2: New Population Generation

```
 1:   for k = 1 to NP do
 2:       Choose β^(k) according to β_i_p
 3:       Decide index^(k) according to g_p
 4:       Initialize Ĉ^(k) (t)
 5:       for i = 1 to n_1 do
 6:           randomly choose a from A
 7:           A = A \{a}
 8:           B = {1, 2, ... m_a}
 9:           for j = 1 to m_a do
10:               randomly choose b from B
11:               B = B \{b}
12:               if index^(k) = 0
13:                   x_ab^(k) = N(μ_ab, σ_ij)
14:               else
15:                   x_ab^(k) = C (μ_ab, σ_ij)
16:               end if
17:               if x_ab^(k) > u_ab(t_0)
18:                   x_ab^(k) = u_ab (t_0)
19:               else if x_ab^(k) < l_abt_0 and x_ab^(k) ≥ l_abt_0/2
20:                   x_ab^(k) = l_abt_0
21:               else if x_ab^(k) < l_abt_0/2
22:                   x_ab^(k) = 0
23:               end if
24:               for t_1 = t_0 to t_0 + y_ab − 1 do
25:                   Ĉ^(k)(t_1) = Ĉ^(k)(t_1) − x_ab^(k) × (t_1 − t_0 + 1)
26:               end for
27:               for t_1 = t_0 + y_ab to T_1 do
28:                   Ĉ^(k)(t_1) = Ĉ^(k)(t_1) − x_ab^(k) × y_ab
29:               end for
30:           end for
31:       end for
32:   end for
```

---

probabilistic models of continuous variables should follow the Gaussian distribution or the Cauchy distribution. The decision is made based on the performances of the two models in the previous generation.

Step 4 (New Population Generation): This process generates a new population of NP feasible individuals by sampling from the probabilistic models built in steps 2 and 3.

Step 5 (Local Search): This process fine-tunes the best-so-far individual by sampling its neighborhood using the Gaussian distribution.

Step 6 (Termination Check): If the termination criterion is satisfied, the best-so-far solution is returned as the result. Otherwise, return to step 2 and start a new generation.

V. SIMULATIONS AND DISCUSSION

A specific example will be given in this section to help better understand how the optimization strategy given in Section IV can get a profitable insurance investment plan regarding the model given in Section III. To validate the effectiveness and efficiency of the proposed model and method, comparisons are made among different investment plans and different EDAs.

TABLE I
SETTINGS OF POLICYHOLDER'S INCOME

Age	Yearly Income
< 20	12000
20 ~ 24	25000
25 ~ 29	40000
30 ~ 34	50000
35 ~ 39	53000
40 ~ 44	50000
45 ~ 49	48000
50 ~ 60	45000
> 60	20000

TABLE II
DETAILED SETTINGS OF THE ENDOWMENT POLICIES

	Payment Period (y_{ij}) and Latest Purchasing Time (q_{ij})					Minimum of the face amount (c_i)	The focus
A	5			10		20,000	Deposit
	75			70			
B	1	5	10	18	25	10,000	Life insurance
	75	65	60	60	55		
C	5			10		10,000	Deposit
	75			70			
D	6	10	15	20		5,000	Life insurance
	65	60	55	55			
E	1		5	10		1,500	Deposit
	80		80	75			

TABLE III
DETAILED SETTINGS OF THE HOSPITALIZATION POLICIES

	Coverage (A_k)	Premium ($\delta_k(t_0)$)
A	critical illness	lowest
B	critical and serious illness	medium
C	critical, serious and general illness	highest

A. Experimental Settings

1) Settings of the Policyholder: We consider the situations that the policyholder has different initial age ($t_0 = 0, 10, 20, 30, 40, 50, 60$) and different initial wealth ($I_0 = 100\,000, 200\,000, 500\,000$). The policyholder's yearly income, as shown in Table I, is set according to an investigation on the average salary of common wage-earners.

Assume that the policyholder is a male without the habit of smoking, and has not been diagnosed with any serious diseases that will affect his health in the future. Therefore, the mortality rate and the incidence rates of diseases with different serious degrees are set in accordance with the regular patterns.

2) Settings of Insurance Policies: In the simulation, five endowment policies and three hospitalization policies are considered. Detailed settings of the policies are tabulated in Tables II and III.

Here, the minimum of face amount c_i is a factor that decides the minimum amount of investment l_{ijt}. For each endowment policy, the minimum face amount is the same while the minimum amount of investment is different: it is the investment amount in accordance with the face amount c_i.

As can be seen, the five endowment policies have different payment periods, latest purchasing time, and face amounts. Except the endowment policy B and D, all the other endowment policies focus on deposit, whose return is thus not related
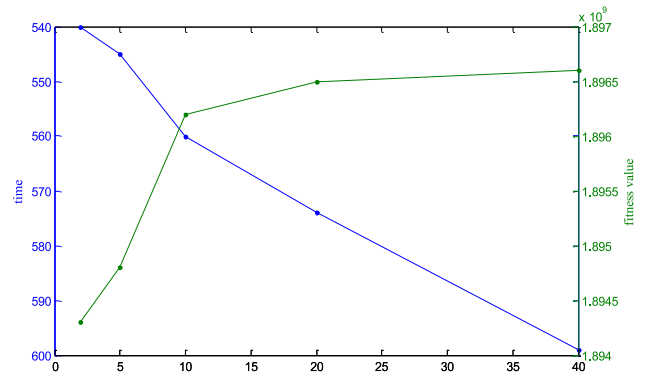

Fig. 3. Comparison between different numbers of individuals during the local search process with situation $t_0 = 0$ and $I_1 = 200\,000$ (the fitness value in the figure is the mean value for 30 runs while the time is the total time for 30 runs).

to the purchasing time. The hospitalization policies have different coverage scopes. To be more exact, from policy A to policy C, the coverage scope expands from critical diseases, critical and serious diseases, to all types of diseases. Their premium also increases accordingly.

3) Settings of Algorithm Parameters: For each experiment, each algorithm is run for 30 independent times with 1000 generations. The population size NP is set to be 300 [34]–[36]. The number of individuals selected for model estimation, NP_{best}, is set as 45% of the whole population, i.e., $NP_{\text{best}} = 135$ (the comparison of different settings of NP_{best} is given in Table IV and shown that $45\% NP$ is most suitable for this optimization problem). The number of individuals generated during the local search is set as 10 (this is the most suitable number with relatively high performance and reasonable running speed according to Fig. 3).

B. Comparison of Different Investment Plans

To illustrate the effectiveness and efficiency of the proposed insurance investment model, we compare its results with the yearly profit obtained by the traditional investment plans that purchase only a single policy (either an endowment policy with a specific payment period or a hospitalization policy). In this way, we can analyze whether it is necessary to consider investing multiple insurance policies.

The comparisons are shown in Fig. 4. With the horizontal axis indicating the time and the vertical axis indicating the amount of yearly profit, the curves in this figure show how the yearly profit of the investment plans generated by different strategies evolve over time. Note that for the proposed model, the results vary in different runs due to the influence of randomness. As such, the figure displays its results over 30 independent runs in the form of box plot.

As can be seen, the yearly profits obtained by the proposed model outperform all the other strategies for different initial ages. Moreover, the result of AEDA is so stable that its box plot is narrowed down to a line.

The effectiveness of the proposed method can also be observed by interpreting the insurance investment plans obtained. The following sections discuss the comparisons

TABLE IV
COMPARISONS BETWEEN DIFFERENT SETTINGS OF NP_{best} WITH SITUATION $T_0 = 10$ AND $I_0 = 200\,000$

|	$NP_{best} = 0.45NP$	$NP_{best} = 0.3NP$	$NP_{best} = 0.35NP$	$NP_{best} = 0.4NP$	$NP_{best} = 0.5NP$	$NP_{best} = 0.55NP$	$NP_{best} = 0.6NP$
Max	1.8976E+09	1.8976E+09	1.8976E+09	1.8976E+09	1.8976E+09	1.8976E+09	1.8976E+09
Median	1.8976E+09	1.8925E+09	1.8925E+09	1.8951E+09	1.8976E+09	1.8973E+09	1.8972E+09
Std	2.3858E+06	3.0979E+06	2.6860E+06	2.8389E+06	1.5307E+06	2.8815E+06	2.7633E+06
p-value	-	2.7751E-02[+]	1.2201E-03[+]	9.5948E-02	2.3321E-01	1.4897E-01	2.1456E-02[+]
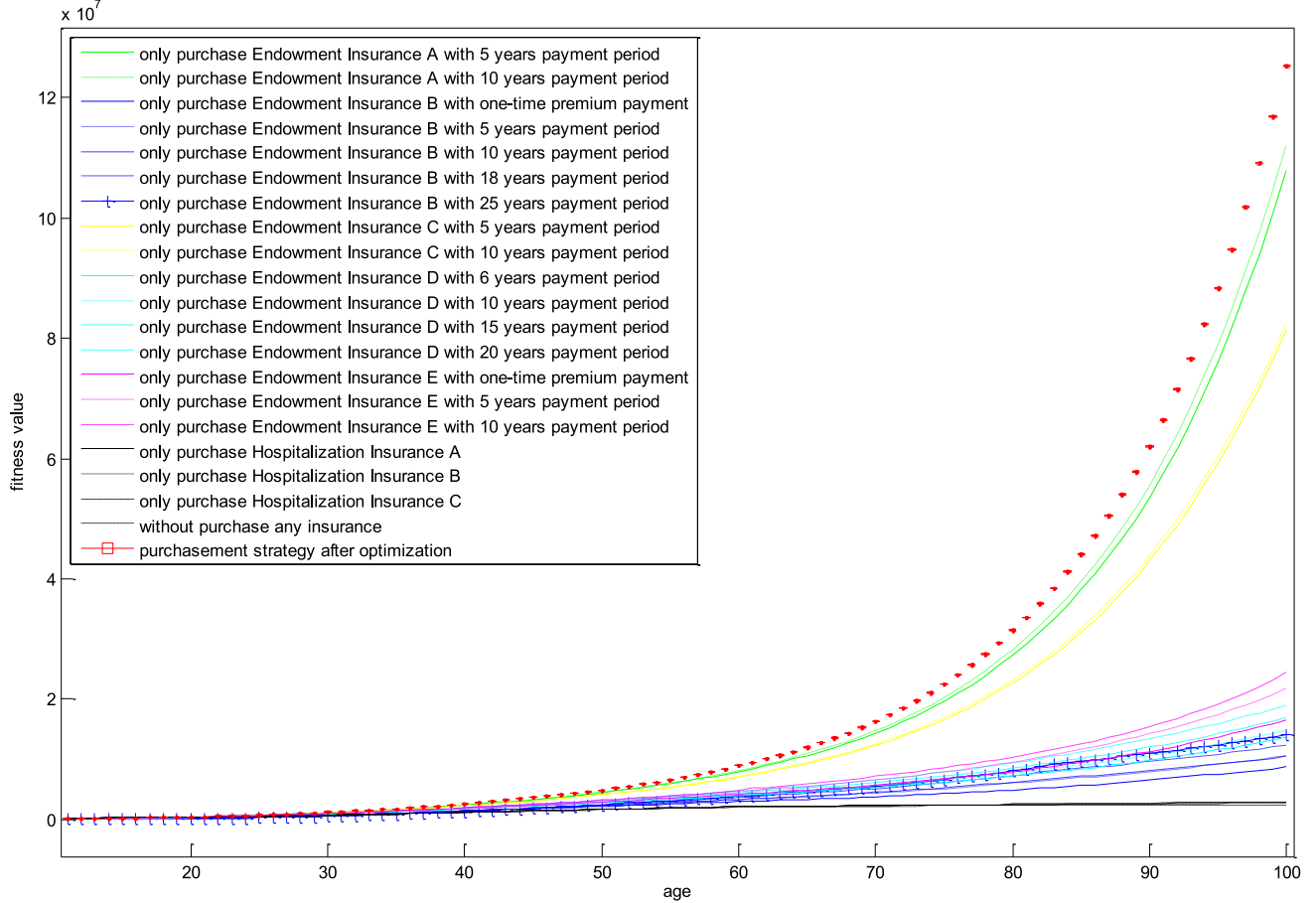
Fig. 4. Comparison of different investment plans for the situation $t_0 = 0$ and $I_1 = 200\,000$.

TABLE V
INVESTMENT PLANS UNDER DIFFERENT INITIAL AGES

	Endowment Policy A		Endowment Policy D	Endowment Policy E	Choice of Hospitalization Policy
Situation	**5-year payment periods**	**10-year payment periods**	**10-year payment periods**	**one-time premium payment**	
$t_0 = 0,\quad I_0 = 200000$	37580	11950	0	0	None
$t_0 = 10,\ I_0 = 200000$	37300	12000	0	1500	B
$t_0 = 20,\ I_0 = 200000$	16700	40000	0	1500	B
$t_0 = 30,\ I_0 = 200000$	26100	53000	0	1500	B
$t_0 = 40,\ I_0 = 200000$	31300	48380	0	1500	B
$t_0 = 50,\ I_0 = 200000$	30710	44960	0	1500	B
$t_0 = 60,\ I_0 = 200000$	35700	19925	0	1600	B

under two situations: 1) different initial age with a fixed initial wealth and 2) different initial wealth with a fixed initial age. By doing so, the advantage of the proposed model can be analyzed in a comprehensive way.

1) Comparison Under Different Initial Ages: The comparison of the results obtained for policyholders with $t_0 = 0, 10, 20, 30, 40, 50, 60$ and $I_0 = 200\,000$ are

given in Table V. Situations for the other initial wealth are similar, which are not shown here due to the page limit.

To obtain the optimal yearly profit, the policyholder at age 0 is suggested to buy endowment policy A with 5- and 10-year payment periods, but not to buy any hospitalization policy. This is reasonable because the probability for

TABLE VI
INVESTMENT PLANS UNDER DIFFERENT INITIAL WEALTH

Situation	Endowment Policy A		Endowment Policy D	Endowment Policy E	Choice of Hospitalization Policy
	5-year payment periods	10-year payment periods	10-year payment periods	one-time premium payment	
$t_0 = 30$, $I_0 = 50000$	0	49087	887	0	A
$t_0 = 30$, $I_0 = 100000$	6100	53000	0	1500	B
$t_0 = 30$, $I_0 = 200000$	26100	53000	0	1500	B
$t_0 = 30$, $I_0 = 500000$	86100	53000	0	1500	B

TABLE VII
COMPARISON OF THE AEDA AND OTHER EDAs WITH DIFFERENT INITIAL AGE

		AEDA	EDA-C	EDA-G	EDA-H	EDA-R	EDA-L
$t_0 = 0$	Max	3.7663E+09	3.7663E+09	3.7508E+09	3.7663E+09	3.7663E+09	3.7663E+09
	Median	3.7663E+09	3.7663E+09	3.6738E+09	3.7663E+09	3.7663E+09	3.7663E+09
	Std	2.1084E-06	1.9476E+03	6.2622E+07	2.7737E+04	2.2361E+01	2.0576E-06
	p-value	-	1.2455E-11$^+$	1.2455E-11$^+$	5.4908E-01	1.2455E-11$^+$	7.9528E-01
$t_0 = 10$	Max	1.8976E+09	1.8966E+09	1.8910E+09	1.8976E+09	1.8976E+09	1.8976E+09
	Median	1.8976E+09	1.8925E+09	1.8482E+09	1.8954E+09	1.8970E+09	1.8951E+09
	Std	2.3858E+06	1.3702E+06	2.5120E+07	1.6795E+06	9.3094E+05	2.0688E+06
	p-value	-	2.1135E-08$^+$	2.8646E-11$^+$	1.2174E-02$^+$	4.1984E-01	8.0161E-03+
$t_0 = 20$	Max	1.3628E+09	1.3627E+09	1.3532E+09	1.3628E+09	1.3628E+09	1.3628E+09
	Median	1.3628E+09	1.3549E+09	1.3287E+09	1.3624E+09	1.3628E+09	1.3627E+09
	Std	3.2016E+05	3.3980E+06	1.5736E+07	6.7314E+06	4.1157E+05	2.4705E+06
	p-value	-	2.0107E-10$^+$	2.8003E-11$^+$	4.5742E-06$^+$	6.0098E-03$^+$	4.6673E-02$^+$
$t_0 = 30$	Max	9.1344E+08	9.1319E+08	9.0386E+08	9.1344E+08	9.1344E+08	9.1344E+08
	Median	9.1344E+08	9.1161E+08	8.9115E+08	9.1318E+08	9.1244E+08	9.1344E+08
	Std	2.0263E+06	3.7981E+06	7.8229E+06	2.4639E+06	8.7003E+05	3.7614E+05
	p-value	-	1.9081E-09$^+$	2.4690E-11$^+$	6.7480E-09$^+$	6.0045E-09$^+$	3.1431E-01
$t_0 = 40$	Max	4.4258E+08	4.4247E+08	4.3991E+08	4.4258E+08	4.4258E+08	4.4258E+08
	Median	4.4258E+08	4.4102E+08	4.3318E+08	4.4258E+08	4.4136E+08	4.4258E+08
	Std	4.0971E+05	3.8133E+06	5.4879E+06	1.1250E+06	7.0785E+05	1.6154E+06
	p-value	-	1.9332E-10$^+$	2.9803E-11$^+$	1.1646E-01	4.4700E-10$^+$	5.9364E-01
$t_0 = 50$	Max	2.0136E+08	2.0136E+08	2.0066E+08	2.0136E+08	2.0136E+08	2.0136E+08
	Median	2.0136E+08	2.0095E+08	1.9641E+08	2.0136E+08	1.9993E+08	2.0136E+08
	Std	1.0129E+06	5.4426E+05	3.1227E+06	8.8405E-02	5.3850E+05	3.9070E+05
	p-value	-	3.8021E-10$^+$	1.1863E-10$^+$	3.8461E-01	2.1562E-10$^+$	8.8053E-01
$t_0 = 60$	Max	6.5234E+07	6.5211E+07	6.4967E+07	6.5234E+07	6.5234E+07	6.5234E+07
	Median	6.5234E+07	6.5047E+07	6.3554E+07	6.5234E+07	6.3903E+07	6.5234E+07
	Std	7.3051E-03	4.7102E+05	6.8100E+05	1.0627E+04	6.3355E+05	3.4327E+05
	p-value	-	2.9991E-11$^+$	2.9991E-11$^+$	5.1938E-03$^+$	2.9953E-11$^+$	6.4097E-01
b/e/w			7/0/0	7/0/0	4/3/0	6/1/0	2/5/0

a 0-year-old baby to get ill is low. Besides, if the policy-holder purchased any hospitalization policy, he has to pay the premium for an extremely long period. At other initial ages, the policyholder is suggested to buy the endowment policy A with 5- and 10-year payment periods and the endowment policy E with one-time premium payment. This is probably because the policyholder has not been diagnosed with any serious diseases and maintained a healthy living style. Therefore, the possibility of his death is relatively low so that this policyholder is not suggested to buy any endowment policy focusing on life insurance. Note that the proposed method also suggests that the optimal amount of investment varies depending on the policyholder's initial age. In terms of the hospitalization policies, the proposed method suggests to choose policy B over the others. The reason is that this policyholder is currently in a good health condition and is unlikely to have general illnesses frequently. Therefore, it is not necessary to buy hospitalization policy C. But the coverage of hospitalization policy A may not be secure enough, because as a common wage-earner, the medical expense will be a large burden for the policyholder once he gets some serious illness.

2) Comparison Under Different Initial Wealth: The comparison for 30-year old policyholders with $I_0 = 50\,000, 100\,000, 200\,000,$ and $500\,000$, are given in Table VI. The situations with other initial wealth are similar and thus not shown here due to the page limit.

It can be seen that for a 30-year old policyholder, when the initial wealth is 50 000, endowment policy A with a 10-year payment period and endowment policy D with a 6-year payment period should be chosen. With the augment of the initial wealth, no matter how much money the policyholder has at the beginning, he should choose endowment policy A with 5- and 10-year payment period and endowment policy E with

TABLE VIII
COMPARISON OF THE AEDA AND OTHER EDAS WITH DIFFERENT INITIAL WEALTH

		EDA-S	EDA-C	EDA-G	EDA-H	EDA-R	EDA-L
$I_0 = 50000$	Max	6.7471E+08	6.7471E+08	6.7019E+08	6.7471E+08	6.7471E+08	6.7471E+08
	Median	6.7471E+08	6.7215E+08	6.5267E+08	6.7215E+08	6.7215E+08	6.7215E+08
	Std	1.1364E+06	8.3258E+05	2.0356E+07	8.8314E+05	1.1931E+06	7.8717E+05
	p-value	-	4.4301E-06$^+$	1.0149E-11$^+$	1.8723E-05$^+$	4.2676E-04$^+$	5.0786E-06$^+$
$I_0 = 100000$	Max	7.6619E+08	7.6615E+08	7.6541E+08	7.6619E+08	7.6619E+08	7.6619E+08
	Median	7.6619E+08	7.6352E+08	7.5279E+08	7.6610E+08	7.6519E+08	7.6619E+08
	Std	2.7165E+06	3.2975E+06	1.1667E+07	1.2903E+06	9.9447E+05	3.1121E+06
	p-value	-	4.2153E-07$^+$	2.0792E-10$^+$	1.4433E-01	3.7876E-06$^+$	9.5613E-01
$I_0 = 200000$	Max	9.1344E+08	9.1319E+08	9.0386E+08	9.1344E+08	9.1344E+08	9.1344E+08
	Median	9.1344E+08	9.1161E+08	8.9115E+08	9.1318E+08	9.1244E+08	9.1344E+08
	Std	2.0263E+06	3.7981E+06	7.8229E+06	2.4639E+06	8.7003E+05	3.7614E+05
	p-value	-	1.9081E-09$^+$	2.4690E-11$^+$	6.7480E-09$^+$	6.0045E-09$^+$	3.1431E-01
$I_0 = 500000$	Max	1.3552E+09	1.3533E+09	1.3516E+09	1.3552E+09	1.3552E+09	1.3552E+09
	Median	1.3552E+09	1.3469E+09	1.3253E+09	1.3542E+09	1.3542E+09	1.3552E+09
	Std	2.3336E+06	3.2177E+06	2.0721E+07	2.3137E+06	4.6603E+05	4.7980E-03
	p-value	-	1.0136E-09$^+$	6.9000E-11$^+$	1.1217E-07$^+$	1.9569E-07$^+$	9.1532E-01
b/e/w			4/0/0	4/0/0	3/1/0	4/0/0	1/3/0

one-time premium payment policy. In terms of the hospitalization policy, when the initial wealth is 50 000, the policyholder is suggested legible to buy policy A because he does not have enough economic strength to buy a more expensive one. As the initial wealth increases, he is suggested to buy policy B. However, even though the initial wealth is 500 000, he is not suggested to buy insurance C. This may be because that we make a hypothesis that the health condition of this policyholder is good so that policy B is sufficient and it is not necessary to buy policy C.

From the results above, it can be found that the proposed data-driven insurance investment strategy considers multi polices outperforms each of the single-policy strategy. Therefore, the necessity of considering different kinds of policies has been proven.

C. Comparisons of Different EDAs

To validate the performance of the proposed AEDA, experiments are performed to compare it with following EDAs.
1) EDA that samples new population with Cauchy distribution only (EDA-C) [49].
2) EDA that samples new population with Gaussian distribution only (EDA-G) [34].
3) EDA that samples half of new population with Cauchy distribution and the other half with Gaussian distribution (EDA-H) [50].
4) EDA that samples the value of β using Cauchy or Gaussian distribution like x_{ij} and then round down to the nearest integer (EDA-R).
5) EDA without local search strategy (EDA-L).

For the sake of fairness, all the six EDAs are executed with the same parameter settings. The maximum value, median value, and standard deviation that each EDA obtained in 30 independent runs are reported for comparison. To examine the significance of the difference between the proposed AEDA and the other five EDAs, the Wilcoxon

rank sum test [51] is performed and the p-value is also reported.

The proposed AEDA and the above five EDAs are first compared with the initial age t_0 varying in the range of {0, 10, 20, 30, 40, 50, 60} and the same initial wealth I_0 fixed at 200 000. The results are tabulated in Table VII. Then the six EDAs are compared with the initial age t_0 fixed at 30 and the initial wealth I_0 varying in the range of {50 000, 100 000, 200 000, 500 000}. The results are shown in Table VIII.

Tables VII and VIII show that AEDA significantly outperforms EDA-C and EDA-D in all cases. It also significantly outperforms EDA-H in 7 out of the 11 cases. Therefore, the effectiveness of the adaptive model selection strategy for choosing sampled distribution has been confirmed. The comparison of AEDA with EDA-R shows that AEDA outperforms EDA-R in 10 out of the 11 cases, which shows the effectiveness of modeling different types of variables using different methods. The comparison of AEDA with EDA-L shows that AEDA outperforms EDA-L for some initial ages and some initial wealth, meaning that it is necessary to employ the local search strategy for some situations.

VI. CONCLUSION

In this paper, a data-driven model is defined for optimization of the insurance investment plan that involves endowment and hospitalization policies. An AEDA is proposed to address the model, offering an individualized way to derive the optimal insurance investment plan that maximizes the yearly profit given the policyholder's initial age and wealth.

The use of evolutionary algorithms in the field of insurance investment optimization is a new attempt. This paper explores the possibility and rationality of such an attempt, and shows that evolutionary algorithms can be an efficient tool for optimizing insurance investment plans. This paper offers a new perspective and an effective way for realizing intelligentization and individualization of insurance investment.

References

[1] N. H. Hakansson, "Optimal investment and consumption strategies under risk, an uncertain lifetime, and insurance," *Int. Econ. Rev.*, vol. 10, no. 3, pp. 443–466, 1967.

[2] S. F. Richard, "Optimal consumption, portfolio and life insurance rules for an uncertain lived individual in a continuous time model," *J. Financ. Econ.*, vol. 2, no. 2, pp. 187–203, 1975.

[3] M. J. Brennan and E. S. Schwartz, "The pricing of equity-linked life insurance policies with an asset value guarantee," *J. Financ. Econ.*, vol. 3, pp. 195–213, 1976.

[4] A. R. Bacinello, "Pricing guaranteed life insurance participating policies with annual premiums and surrender option," *North Amer. Actuarial J.*, vol. 7, no. 3, pp. 1–17, 2003.

[5] J. Li and A. Szimayer, "The uncertain mortality intensity framework: Pricing and hedging unit-linked life insurance contracts," *Insur. Math. Econ.*, vol. 49, no. 3, pp. 471–486, 2010.

[6] A. R. Bacinello and F. Ortu, "Pricing equity-linked life insurance with endogenous minimum guarantees," *Insur. Math. Econ.*, vol. 12, no. 3, pp. 245–257, 1993.

[7] K. K. Aase and S.-A. Persson, "Pricing of unit-linked life insurance policies," *Scandinavian Actuarial J.*, vol. 1994, no. 1, pp. 26–52, 1994.

[8] A. R. Bacinello and F. Ortu, "Fixed income linked life insurance policies with minimum guarantees: Pricing models and numerical results," *Eur. J. Oper. Res.*, vol. 91, no. 2, pp. 235–249, 1996.

[9] H. Ma, C. Song, and Y. Liu, "A review of empirical literature of non-comparative international entrepreneurship and future prospects," *Foreign Econ. Manag.*, vol. 38, no. 2, pp. 16–32, 2016.

[10] S. R. Pliska and J. Ye, "Optimal life insurance purchase and consumption/investment under uncertain lifetime," *J. Bank. Financ.*, vol. 31, no. 5, pp. 1307–1319, 2007.

[11] J. Ye, "Optimal life insurance, consumption and portfolio under uncertainty: Martingale methods," in *Proc. Amer. Control Conf.*, 2007, pp. 1103–1109.

[12] I. Duarte, A. A. Pinto, and S. R. Pliska, "Optimal life insurance purchase, consumption and investment on a financial market with multi-dimensional diffusive terms," *Optimzation*, vol. 63, no. 11, pp. 1737–1760, 2011.

[13] I. Duarte, D. Pinheiro, A. A. Pinto, and S. R. Pliska, "An overview of optimal life insurance purchase, consumption and investment problems," *Dyn. Games Sci. I*, vol. 1, pp. 271–286, Jan. 2011.

[14] A. S. Mousa, D. Pinheiro, and A. A. Pinto, "Optimal life-insurance selection and purchase within a market of several life-insurance providers," *Insur. Math. Econ.*, vol. 67, pp. 133–141, Mar. 2016.

[15] M. Kwak, Y. H. Shin, and U. J. Choi, "Optimal investment and consumption decision of a family with life insurance," *Insur. Math. Econ.*, vol. 48, no. 2, pp. 176–188, 2011.

[16] H. Huang, M. A. Milevsky, and J. Wang, "Portfolio choice and life insurance: The CRRA case," *J. Risk Insur.*, vol. 75, no. 4, pp. 847–872, 2008.

[17] E. Bengoetxea, P. Larrañaga, I. Bloch, and A. Perchant, "Estimation of distribution algorithms," *Genet. Algorithms Evol. Comput.*, vol. 64, no. 5, pp. 454–468, 2001.

[18] Q. Yang *et al.*, "Segment-based predominant learning swarm optimizer for large-scale optimization," *IEEE Trans. Cybern.*, vol. 47, no. 9, pp. 2896–2910, Sep. 2017.

[19] Q. Liu *et al.*, "Benchmarking stochastic algorithms for global optimization problems by visualizing confidence intervals," *IEEE Trans. Cybern.*, vol. 47, no. 9, pp. 2924–2937, Sep. 2017.

[20] X. Wen *et al.*, "A maximal clique based multiobjective evolutionary algorithm for overlapping community detection," *IEEE Trans. Evol. Comput.*, vol. 21, no. 3, pp. 363–377, Jun. 2017.

[21] Q. Yang *et al.*, "Adaptive multimodal continuous ant colony optimization," *IEEE Trans. Evol. Comput.*, vol. 21, no. 2, pp. 191–205, Apr. 2017.

[22] Q. Yang *et al.*, "A level-based learning swarm optimizer for large scale optimization," *IEEE Trans. Evol. Comput.*, to be published.

[23] M. S. Dorfman, *Introduction to Risk Management and Insurance*. Upper Saddle River, NJ, USA: Prentice-Hall, 2004.

[24] R. B. Goyette, "Hospitalization insurance," *Lunión Médicale Du Canada*, vol. 86, no. 12, pp. 1396–1400, 1957.

[25] S. Baluja, "Population-based incremental learning. A method for integrating genetic search based function optimization and competitive learning," Carnegie Mellon Univ., Pittsburgh, PA, USA, Rep. CMU-CS-94-163, 1994.

[26] H. Mühlenbein, "The equation for response to selection and its use for prediction," *Evol. Comput.*, vol. 5, no. 3, pp. 303–346, 1997.

[27] G. R. Harik, F. G. Lobo, and D. E. Goldberg, "The compact genetic algorithm," *IEEE Trans. Evol. Comput.*, vol. 3, no. 4, pp. 287–297, Nov. 1999.

[28] J. S. de Bonet, C. L Isbell, and P. Viola, "MIMIC: Finding optima by estimating probability densities," in *Advances in Neural Information Processing Systems*. Cambridge, MA, USA: MIT Press, 1996, pp. 424–430.

[29] S. Baluja and S. Davies, "Using optimal dependency-trees for combinatorial optimization: Learning the structure of the search space," in *Proc. Int. Conf. Mach. Learn.*, 1997, pp. 30–38.

[30] M. Pelikan and H. Muehlenbein, "The bivariate marginal distribution algorithm," in *Advances in Soft Computing*. London, U.K.: Springer, 1999, pp. 521–535.

[31] H. Muhlenbein and T. Mahnig, "Convergence theory and applications of the factorized distribution algorithm," *J. Comput. Inf. Technol.*, vol. 7, no. 1, pp. 19–32, 1999.

[32] G. Harik, "Linkage learning via probabilistic modeling in the ECGA," Univ. Illinois Urbana–Champaign, Champaign, IL, USA, Rep. 99010, 1999.

[33] M. Pelikan, D. E. Goldberg, and E. Cantú-Paz, "BOA: The Bayesian optimization algorithm," in *Proc. Conf. Genet. Evol. Comput.*, 1999, pp. 525–532.

[34] M. Sebag and A. Ducoulombier, "Extending population-based incremental learning to continuous search spaces," in *Proc. Int. Conf. Parallel Problem Solving Nat.*, 1998, pp. 418–427.

[35] P. Larrañaga, R. Etxeberria, J. A. Lozano, and J. M. Peña, "Optimization in continuous domains by learning and simulation of Gaussian networks," in *Proc. Genet. Evol. Comput. Conf.*, 2000, pp. 201–204.

[36] S. Rudolf and M. Köppen, "Stochastic hill climbing by vectors of normal distributions," in *Proc. 1st Online Workshop Soft Comput. (WSC1)*, 1996, pp. 60–70.

[37] R. Etxeberria, J. A. Lozano, and J. M. Peña, "Optimization by learning and simulation of Bayesian and Gaussian networks," Univ. Basque Country, Leioa, Spain, Rep. EHU-KZAA-IK-4/99, 1999.

[38] P. A. N. Bosman and D. Thierens, "Expanding from discrete to continuous estimation of distribution algorithms: The ID𝔼A," in *Proc. Int. Conf. Parallel Problem Solving From Nat.*, 2000, pp. 767–776.

[39] J. Ceberio, E. Irurozki, A. Mendiburu, and J. A. Lozano, "A distance-based ranking model estimation of distribution algorithm for the flow-shop scheduling problem," *IEEE Trans. Evol. Comput.*, vol. 18, no. 2, pp. 286–300, Apr. 2014.

[40] X. Li, S. Mabu, and K. Hirasawa, "A novel graph-based estimation of the distribution algorithm and its extension using reinforcement learning," *IEEE Trans. Evol. Comput.*, vol. 18, no. 1, pp. 98–113, Feb. 2014.

[41] W. Dong, T. Chen, P. Tino, and X. Yao, "Scaling up estimation of distribution algorithms for continuous optimization," *IEEE Trans. Evol. Comput.*, vol. 17, no. 6, pp. 797–822, Dec. 2013.

[42] R. Santana, P. Larrañaga, and J. A. Lozano, "Protein folding in simplified models with estimation of distribution algorithms," *IEEE Trans. Evol. Comput.*, vol. 12, no. 4, pp. 418–438, Aug. 2008.

[43] Y. Hasegawa and H. Iba, "Latent variable model for estimation of distribution algorithm based on a probabilistic context-free grammar," *IEEE Trans. Evol. Comput.*, vol. 13, no. 4, pp. 858–878, Aug. 2009.

[44] Y. Chen *et al.*, "Personalized search inspired fast interactive estimation of distribution algorithm and its application," *IEEE Trans. Evol. Comput.*, vol. 21, no. 4, pp. 588–600, Aug. 2017.

[45] Q. Yang *et al.*, "Multimodal estimation of distribution algorithms," *IEEE Trans. Cybern.*, vol. 47, no. 3, pp. 636–650, Mar. 2017.

[46] P. Yang, K. Tang, and X. Lu, "Improving estimation of distribution algorithm on multimodal problems by detecting promising areas," *IEEE Trans. Cybern.*, vol. 45, no. 8, pp. 1438–1449, Aug. 2015.

[47] V. A. Shim, K. C. Tan, and C. Y. Cheong, "A hybrid estimation of distribution algorithm with decomposition for solving the multiobjective multiple traveling salesman problem," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 42, no. 5, pp. 682–691, Sep. 2012.

[48] (2017). *Savings—AIA Hong Kong*. [Onilne]. Available: http://www.aia.com.hk/en/our-products /savings.html

[49] P. Pošík, "BBOB-benchmarking a simple estimation of distribution algorithm with Cauchy distribution," in *Proc. Genet. Evol. Comput. Conf. (GECCO)*, Montreal, QC, Canada, Jul. 2009, pp. 2309–2314.

[50] N. Luo and F. Qian, "Estimation of distribution algorithm sampling under Gaussian and Cauchy distribution in continuous domain," in *Proc. IEEE Int. Conf. Control Autom.*, 2010, pp. 1716–1720.

[51] W. Haynes, *Wilcoxon Rank Sum Test*. New York, NY, USA: Springer, 2013, pp. 2354–2355.

Wen Shi (S'16) received the bachelor's degree from Sun Yat-sen University, Guangzhou, China, in 2016. She is currently pursuing the Ph.D. degree with the School of Computer Science and Engineering, South China University of Technology, Guangzhou.

Her current research interests include evolutionary computation algorithms and their applications on financial problems, differential evolution, particle swarm optimization, and estimation of distribution algorithms and their applications on financial problems, such as insurance investment optimization problem.

Wei-Neng Chen (S'07–M'12–SM'17) received the bachelor's and Ph.D. degrees from Sun Yat-sen University, Guangzhou, China, in 2006 and 2012, respectively.

He is currently a Professor with the School of Computer Science and Engineering, South China University of Technology, Guangzhou. His current research interests include swarm intelligence algorithms and their applications on cloud computing, operations research, and software engineering. He has published over 70 papers in international journals and conferences, including over 20 papers in IEEE TRANSACTIONS journals.

Dr. Chen was a recipient of the IEEE Computational Intelligence Society Outstanding Dissertation Award in 2016 for his Doctoral Thesis, and the National Science Fund for Excellent Young Scholars in 2016.

Ying Lin (M'12) received the Ph.D. degree in computer applied technology from Sun Yat-sen University, Guangzhou, China, in 2012.

She is currently an Assistant Professor with the Department of Psychology, Sun Yat-sen University. Her current research interests include computational intelligence and its applications in network analysis and cognitive diagnosis.

Tianlong Gu received the M.Eng. degree from Xidian University, Xi'an, China, in 1987 and the Ph.D. degree from Zhejiang University, Hangzhou, China, in 1996.

From 1998 to 2002, he was a Research Fellow with the School of Electrical and Computer Engineering, Curtin University of Technology, Perth, WA, Australia, and a Post-Doctoral Fellow with the School of Engineering, Murdoch University, Perth. He is currently a Professor with the School of Computer Science and Engineering, Guilin University of Electronic Technology, Guilin, China. His current research interests include formal methods, data and knowledge engineering, software engineering, and information security protocol.

Sam Kwong (M'93–SM'04–F'14) received the B.Sc. degree in electrical engineering from the State University of New York at Buffalo, Buffalo, NY, USA, in 1983, the M.A.Sc. degree in electrical engineering from the University of Waterloo, Waterloo, ON, Canada, in 1985, and the Ph.D. degree from the University of Hagen, Hagen, Germany, in 1996.

From 1985 to 1987, he was a Diagnostic Engineer with Control Data Canada, Mississauga, ON, Canada, where he designed the diagnostic software to detect the manufacture faults of the very large scale integration chips in the Cyber 430 machine. He joined the Bell Northern Research Canada, Ottawa, ON, Canada, as a Member of Scientific Staff. In 1990, he joined the City University of Hong Kong, Hong Kong, as a Lecturer with the Department of Electronic Engineering, where he is currently a Professor with the Department of Computer Science. His current research interests include pattern recognition, evolutionary computations, and video analytics.

Prof. Kwong has been the Vice President for IEEE Systems, Man and Cybernetics for conferences and meetings since 2014. He is also appointed as an IEEE Distinguished Lecturer for IEEE SMC Society from 2017.

Jun Zhang (M'02–SM'08–F'17) received the Ph.D. degree in electrical engineering from the City University of Hong Kong, Hong Kong, in 2002.

From 2004 to 2016, he was a Professor with Sun Yat-sen University, Guangzhou, China. Since 2016, he has been with South China University of Technology, Guangzhou, where he is currently a Cheung Kong Chair Professor. His current research interests include computational intelligence, cloud computing, big data, high performance computing, data mining, wireless sensor networks, operations research, and power electronic circuits. He has authored seven research books and book chapters, and over 100 technical papers in the above areas.

Prof. Zhang was a recipient of the China National Funds for Distinguished Young Scientists from the National Natural Science Foundation of China in 2011 and the First-Grade Award in Natural Science Research from the Ministry of Education, China, in 2009. He is currently an Associate Editor of the IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION, the IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS, and the IEEE TRANSACTIONS ON CYBERNETICS. He is the Founding and Current Chair of the IEEE Guangzhou Subsection and IEEE Beijing (Guangzhou) Section Computational Intelligence Society Chapters. He is the Founding and Current Chair of the ACM Guangzhou Chapter. He is a fellow of Institute of Electrical and Electronics Engineers.