Research on DNA Computing Methods of Optimization Problems on Weighted Graph

Liu Jieyu

(Sichuan Tianyi University at Chengdu,Chengdu,Sichuan 618200)

Keywords: DNA computing；Algorithm；Optimization problems；Empowerment

Abstract. The paper researched the DNA computing method of classical optimization problems on weighted graph, improved the method of Weight-coding that belongs to the original model of DNA, raised some new methods of DNA encoding and DNA computing. Generally speaking, through the Relative length graph design that belongs to an Undirected weighted graph , given DNA coding and DNA computing a new method by Traveling salesman problems and Chinese mail carrier problems that based on the General line graph of an Undirected weighted graph. And the same time, through the Inverse optimal alignment of DNA sequence given the DNA cording and DNA computing a new method. The method based on the Optimal alignment which one of the Minimum spanning tree problems. Thought the Polynomial transformation about the Vertex cover problems that was the way to improve the Polynomial transformation form Vertex cover problem to Hamiltonian path. The new computing method improved the ability to express and deal with figures during DNA computing.

1. Introduction

DNA computing is a new paradigm of parallel computation based on the interaction between DNA molecules,which develops theoretical computer science by means of molecular biology techniques. The procedure of DNA computing uses DNA molecules as information storage media, and uses the basic techniques of molecular biology as computational operators for copying, synthesizing, separating, sorting, and concatenating information, which solves science problems based on the complement attribute between the bases of different DNA strands. The procedure can be divided into three phases: encoding, computation(molecular biology experiments), and the abstraction of solutions, where the encoding phase maps the problem we want to solve to a set of DNA molecules that can be used to compute by means of molecular biology techniques. The quality of codes directly affects the speed and efficiency of computation.

Currently, the research on DNA computing has made a great progress in theory or experiment methods. Adleman[1] presented a DNA computing model of solving the Hamiltonian path problem in 1994. Lipton[2] abstracted a parallel model of molecular computation to solve the 3-SAT problem on the basis of Adleman's experiment in 1995. Ouyang et al[3] solved the maximal clique problem in 1997. These previous researches on DNA computing do not need to consider the method of encoding weights in DNA strands. With the development of DNA computing techniques, it can be used to solve increasing optimization problems. Therefore, the method of encoding weights in DNA strand is an important issue toward expanding the capability of DNA computing to solve optimization problems. There exist previous works to represent weights in DNA molecules. Narayanan et al[4] presented a method of encoding weights by the lengths of DNA strands. Shin et al[5] proposed a method of encoding weights by the number of hydrogen bonds in fixed length DNA strand. Yamamoto et al[6] proposed a method of encoding weights by the concentrations of DNA strands. Lee et al[7] proposed a method of encoding weights by the melting temperatures of fixed-length DNA strands. These encoding methods successfully solve some instances, but there are certain limitations. So we need to do further research on DNA encoding method.

This paper discusses DNA computing methods for the Chinese postman problem(CPP),the traveling salesman problem(TSP), the minimum spanning tree problem (MST), and the vertex cover problem (VCP) . By designing DNA encoding methods for certain optimization problems, it takes full advantage of different DNA sequences to en-code weights. This work improves the capability of

representing and dealing with data and expands the extent of solving optimization problems in DNA computing.

2. Preliminaries to DNA Computing

DNA (deoxyribonucleic acid) is a kind of polymer that is composed of many monomers (deoxyribonucleotide, briefly nucleotide). There are two modes of connecting DNA monomers to a polymer[8]: (1) 5-phosphate group of one nucleotide combines with 3-hydroxyl group of another nucleotide through phosphodiester bond. Thus, DNA monomers can combine with each other to form a DNA strand. (2) The base of one nucleotide attracts the base of another nucleotide through hydrogen bonds based on the principle of Watson Crick complementary bases. Thus, two DNA strands of reverse complement can combine with each other to form a DNA double-strand.

When DNA monomers combine with each other to form a DNA strand, the combination has cetain orientation[9-10].Usually, the DNA strand with 5 free-end being located at the most left end and 3 free-end being located at the most right end is marked with $5' - X_1X_2 \ldots X_n - 3'$, and the DNA strand with 3 free-end being located at the most left end and 5 free-end being located at the most right end is marked with $3' - X_1X_2 \ldots X_n - 5'$, where X_i denotes the *i* th base in DNA strand. When DNA strands combine with each other to form DNA double-strand through hydrogen bonds, the combination is subject to the principle of Watson-Crick complementary bases, that is, *A* matches *T* and *C* matches *G*.Note that, a DNA double-strand is composed of two antiparallel strands subject to the principle of Watson-Crick complementary bases.

In order to easily describe the methods, the following representation[11-12] is used in this paper: $5'- X_1X_2 \ldots X_n -3'$ is written as $X_1X_2 \ldots X_n$, and $3'- X_1X_2 \ldots X_n -5'$ is written as - $X_1X_2 \ldots X_n$. For any DNA strand *s*, its reversal - *s* is called the reverse strand of *s*. Let *h(s)* denote the mapping function from each base in *s* to its complementary base, that is, *h(A) = T, h(G) = C,h(C) = G,h(T) = A*. The DNA strand *h (s*) is called the complement strand of s, and its reversal - *h(s*) is called the reverse complement of s. Any DNA strand *s* can combine with its reverse complement - *h(s*) to form a DNA double-strand through hydrogen bonds.

3. DNA Computing Method Based on Relative Length Graph for TSP

For a weighted, connected and undirected graph $G = (V, E), vi \in V, eij \in E, 1 \le i, j \le n$, where the weight on edge e_{ij} is w_{ij}, the traveling salesman problem (TSP) is to find a circle with minimum sum of weights that starts from a certain vertex, ends at the same vertex, and passes through every other vertices once and only once. For any instance of the traveling salesman problem, if some weights are very small or very large, the previous encoding method[4] cannot work well. We devise a new DNA encoding method and DNA algorithm based on relative length graph to deal with it. The method encodes weights by the order numbers of weights instead of weights themselves, so it can directly deal with weights of either real numbers or integers, and the obtained solution is not proportional to the length of DNA strand. This makes it can deal with weights in a very broad range.

3.1 Concepts of Order Number of Weight and Relative Length Graph

Definition 1. For a w eighted and undirected graph $G = (V, E), vi \in V, eij \in E, 1 \le i, j \le n$, where the weight on edge e_{ij} is w_{ij}, all the weights are sorted in an increasing order, and the equal weights are at the same position. Thus, all the weights are divided into *p* groups according to their position, which are numbered from 1 to *p* , respeetively. T he number of a weight is called the order number of the weight[13].

Definition 2. For a weighted and undirected graph $G = (V, E), vi \in V, eij \in E, 1 \le i, j \le n$, where the weight on edge e_{ij} is w_{ij}, we remark w_{ij} with w_{ij}, *k* if the order number of *wij* is k. For each remarked weight w_{ij},*k*, we add *k*- 1 nodes on edge e_{ij}. The obtained graph *G?* is called the relative length graph of *G*[13].

According to the definition of relative length graph, the edge e_{ij} is divided into *k* segments if the order number of w_{ij} is *k*. The bigger the order number, the more the segments of the edge. That is, the

number of segments of an edge represents the relative length of the edge. Note that the number of segments isn't directly proportional to the value of weight. Consider the edges with weights 2, 1000,1002,the numbers of segments are 1,2,3,respectively.

3.2 DNA Encoding Method Based on Relative Length Graph

For a weighted and undirected graph $G=(V,E), vi \in V, eij \in E, 1 \le i, j \le n$, where the weight on edge *eij* is $wij, wij \ge 0$, the DNA encoding method based on relative length graph for the traveling salesman problem is given as follows:

(1)Remark each weight w_{ij}. with w_{ij}, *k* according to the order number *k* of w_{ij} .

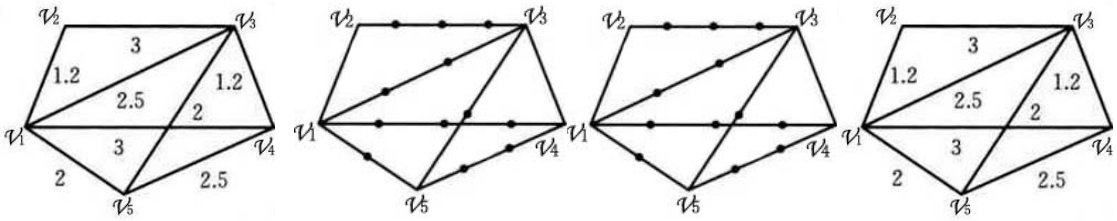
(2) For each remarked weight $w^{ij,k}$, we add *k*-1 nodes on edge e_{ij}. The added nodes are respectively marked with $v'_{e_{ij,1}}, v'_{e_{ij,2}}, \ldots, v'_{e_{ij,k-1}}$.T h e obtained graph is marked with G' .

(3) For each vertex in G', use DNA strand s^i of length $2c(c \in z, c \ge 5)$ $v'i$ to encode it.

(4)For each edge $e'ij$ in G', use DNA strand $sij = -h(si''sj')$, to encode it, where si'' denotes the last half of *si*, sj' denotes the first half of *sj*, - h(s) denotes the reverse complement of *s*.

Thus, when the vertex codes and the edge codes are mixed together, they can combine with each other to form DNA double strands based on the principle of Watson-Crick complementary bases. This method of encoding weights in DNA strands is called the Relative Length Method (RLM).

Take the graph *G* shown in Fig.1(a) as an example. All the weights are sorted in an increasing order, that is, 1.2,2, 2.5,3,so the order numbers of weights 1.2,2,2.5,3 are 1,2,3,4,respectively. Thus, the weights *w*12, *w*13, *w*14, *w*15, *w*23, *w*34, *w*35, *w*45 are remarked with *w12,1*, *w13,3*, *w14,4*, *w15,2*, *w23,4*, *w34,1*, *w35,2*, *w45,3*, respectively. For each remarked weight *wij,k*, we add *k*- 1 nodes on edge *eij* . For example, we add two nodes on *e13* since *w13* is remarked with *w13,3;* we add one node on e15 since *w15* is remarked with *w15,2*, and so on. T he obtained graph *G'* is the relative length graph of *G*, as shown in Fig. 1 (b) .



(a) (b)

Fig. 1 A weighted graph *G* & its relative length graph *G'*

For each vertex *v'i* in Fig. 1(b) , use DNA strand s^i of length 10 (select *c* = 5) to encode it.Here, the added nodes are viewed as vertices. For each edge *e'ij* in Fig. 1(b), use DNA strands *sij= -h(s"is'j)* to encode it. Thus, with the help of the property of reverse complementation between the vertex codes and the edge codes, DNA double strands extend continually to form various random paths including the optimal solution.

3.2 DNA Algorithm Based on Relative Length Graph

Given a weighted and undirected graph $G=(V,E), vi \in V, eij \in E, 1 \le i, j \le n$, where the weight on edge *eij is wij, wij ≧0*. Suppose that *v1* is the original vertex just as well. The DNA algorithm based on relative length graph for the traveling salesman problem is given as follows:

(1) Construct the relative length graph $G'=(V', E')$ of the graph *G* = (V,E).

(2)For the relative length graph G', use DNA strand *si* of length $2c(c \in Z, c \ge 5)$ to encode vertex *v'i*, and use DNA strand *sij= - h(s"is'j)* to encode edge *e'ij* .

(3) Merge:DNA strands *si* and *sij* are mixed together in a single ligation reaction. Based on the principle of Watson-Crick complementary bases, randomly form various DNA double strands.

(4)Denature: DNA double strands are changed into DNA strands by heating.

(5) *B(N, s1)*: Keep all the DNA strands with $5'$- end being s1 , that is, keep all the random paths with the original point being *v*'1 .

(6) *(N, s^1)*: Keep all the DNA molecules with $3'$ -end being s1, that is, keep all the random paths

with the destination point being *v ’1* .

Note that, by the steps (5) and (6) , we get all the DNA molecules with 5’-end and 3 ’-end being s 1.

(7) (*N, si*) : For each vertex *v* ’i except v‘ 1 , keep all the DNA strands with s^i, that is, keep all the paths with *v ’i.*

(8) Separate the shortest path by means of gel electrophoresis.

(9) Sequence the shortest path to determinate its nucleotide sequence, which corresponds to the optimal solution.

3.4 Comparison with Previous Methods

Narayanan et al[4] proposed a length-based DNA encoding method for the shortest path problem and Han et al[14] gave a length-based DNA encoding method for the traveling salesman problem, which encode weights according to the actual length of distance. The relative length method (RLM) is complementary to the previous methods. (1) The previous methods[4,14] can only deal with weights of integer, and weights cannot be very small or very large. RLM can deal with weights of either real number or integer, even very small or very large weight. This is because that,in relative length method, the DNA strands used to encode very large weights are shortened and the DNA strands used to encode very small weights are enlarged. Consider the weights 300,1,1000. The order numbers are 2,1,3, respectively, so the DNA strands used to encode them are with lengths of 20,10,30. (2) Tn the previous methods, when the difference between the lengths of two paths is very small, such as 100 and 101, gel electrophoresis cannot distinguish them[4,14] . This problem can be easily conquered in RLM because with the help of relative length graph, the difference between the lengths of different paths is always above or equal to 2*c*, *c* ≧5. (3) In the previous methods, the length of each path is directly proportional to that of DNA strand. Thus, the weights that can be encoded by the previous methods fall in a narrow range. But the solution obtained in RLM isn’t proportional to the length of DNA strand, which makes RLM can encode weights in a broad range.

Note that, constructing the relative length graph of a weighted graph needs to add $\sum_{i=1}^{e-1} i = \frac{e(e-1)}{2}$ nodes in the worst case, where *e* denotes the num-ber of edges in the given graph. Compared with the original graph, the scale of the relative length graph increases by O (n^2). Therefore, only when there exist very small or very large weights in the given graph or the difference between the lengths of tw o paths is very small, that is, in the case of the length-based DNA encoding method cannot work well, the relative length method is used to encode weights. So the DNA encoding method based on relative length graph is complementary to the length-based DNA encoding method[4,14].

4. DNA Computing Method Based on General Line Graph for CPP

For a weighted, connected and undirected graph $C = (V, E), vi \in V,\ ej \in E, 1 \le i \le n, 1 \le j \le m$, where the weight on edge e_i is *wj*, the Chinese postman problem (CPP) is to find a circle with minimum sum of weights that starts from a certain vertex *vi*, ends at *vi*, and passes through each edge at least once. For the Chinese postman problem, this section first proposes the concept of general line graph, and then devises a DNA encoding method and DNA algorithm based on general line edg e.

4.1 Definition and Construction of General Line Graph

Definition3. For an undirected graph $G = (V, E), vi \in V,\ ej \in E, 1 \le i \le n, 1 \le j \le m$, a mapping function *f* is constructed to satisfy: For each edge $ej \in E$, only one vertex *v’j* in *G* ’ satisfies *f (ej) = v’j;* for each vertex *v’j* in *G* ‘, only one edge *ej* in *G* satisfies

f - 1 (v’j) = ej. If *ei* and *ej* in *G* are adjacent, then add an undirected edge between *v* ’ *i* and *v’j* in *G* ’; if *vi* in *G* is an odd degree vertex, then add at most one selloop to each of the vertices which are mapped from the edges linked to *vi*. The undirected graph obtained by the mapping fro m edge to vertex is called the general line graph of *G*[15].

Take the weighted and undirected graph *G* shown in Fig. 2(a) as an example. T he procedure of mapping from edges to vertices is as follows:（1）Map the edges *e1, e2,…,e8* to the vertices *v’1,*

v'2,…, v '8; (2) Link the vertices in G' : Since the edge *e1* is adjacent to *e2, e8, e7*, the vertex *v'1* is linked to *v'2, v' 8, v'7*; since the edge *e2* is adjacent to *e3, e8, e1*, the vertex *v2* is linked to *v'3, v'8, v' 1* , and so on. (3) Add the self-loops in *G'*: Since *v2* and *v6* are odd degree vertices and the edges linked to *v2* and *v 6* are *e1, e2, e8, e5, e6*, the self loops are respectively added to *v'1, v' 2, v' 8, v' 5, v'6*. The construtted general line graph *G'* is shown in Fig. 2(b) .

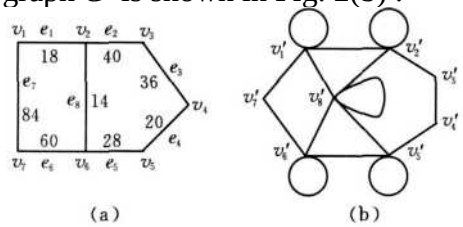


Fig. 2 A weighted and undirected graph *G* and its general line graph *G'*

By means of mapping from edges to vertices,the problem of searching for the shortest circle that passes through each edge at least once is converted into that of searching for the shortest circle that passes through each vertex at least once. Note that the verte?-edge sequence corresponding to the shortest circle isn' t only one. For example, the shortest circle in Fig.*2(a) is v1e1 v2e2v3e3v4e4v5e5v6e8v2e'8v6e6v7e7v1 or v1e1v2e'8 v6e8v2 e2v3e3v4e4v5e5v6e6v7e7v1*, where *e'8* denotes the reversal of *e8*.

4. 2 DNA Encoding Method Based on General Line Graph

Given any weighted, connected and undirected graph $C=(V,E),vi\in V,\ ej\in E,1\le i\le n,1\le j\le m$, where the weight on *ej* is *wj, wj* ≧ *0, wj* ∈ *Z* .In order to easily describe the method, suppose that all the weights in graph *G* are even. If there exists odd weights,all the weights multiplied by 2.In this case,the obtained optimal solution needs to be

divided by 2. T he DNA encoding method based on general line graph for the Chinese postman problem is given as follows:

(1) Construct the general line graph *G'* of the given graph *G* by means of mapping from edges to vertices.

(2) For each vertex *v'j* in *G'* , use DNA stand *sj* of length *wj* to encode it, where *wj* is the weight on edge *ej* corresponding to vertex *v'j* .

(3) For each edge *e'ij = (v'j ,v'j)* in *G'* , use DNA strand *sij = - h(s"is'j)* to encode it, where *s"i* denotes the last half of *sj*, and *-h(s*） denotes the first half of *sj* , and *h(s)* denotes the reverse complement of *s*.

Take the weighted and undirected graph *G* shown in Fig. 2(a) as an example. The general line graph *G'* is first constructed, as shown in Fig. 2(b). For the vertices *v'1, v'2,…v'8* in *G'* , select DNA strand *s1, s2,…,s8* of lengths 18, 40, 36, 20, 28, 60, 84, 14 to encode them, respectively. For each edge *e'12= (v'1, v'2*) in *G* , use DNA strand *sij = - h(s"1s'2* ） to encode it. For example, the code of edge *e' 12= (v '1, v' 2)* is *sn= - h(s"1 s'2)*, whose length is 18/ 2+ 40/2= 29. Thus, with the help of the property of reverse complementation between the vertex codes and the edge codes, DNA double strands extend continually to form various random paths including the optimal solution.

4. 3 DNA Algorithm Based on General Line Graph

Given a weighted, connected and undirected graph $G=(V,E),vi\in V,\ ej\in E,1\le i\le n,1\le j\le m$, where the weight on *ej* is *wj, wj* ≧ *0, wj* ∈ *Z.* Suppose that *v'1* is the original vertex in the general line graph *G'* just as well. DNA algorithm based on general line graph for the Chinese postman problem is given as follows:

(1)Construct the general line graph *G'* of the given graph *G* by means of mapping from edges to vertices.

(2)For each vertex *v'j* in *G'* , use DNA stand *sj* of length *w j* to encode it. For each edge *e'ij = (v' i,v' j)* in *G'* , use DNA strand *sij = -h(s"is 'j)* to encode it, whose length is $|wi|/2+|wj|/2.$

(3)Merge: DNA strands *sj* and *sij* are mixed together in a single ligation reaction. Based on the principle of Watsom-Crick complementary bases, randomly form various DNA double strands.

(4)Denature: DNA double strands are changed into DNA strands by heating.

(5)*B (N, s1*): Keep all the DNA molecules with 5'-end being s1, that is, keep all the random paths with the original point being *v* 1 .

(6)*E (N, s1*): Keep all the DNA molecules with *3'* -end being *s1,* that is, keep all the random paths with the destination point being *v1* .

(7)+ *(N, sj)* : For each vertex *v'j (2 ≤j ≤m)* in *G* , keep all the paths with *v 'j.*

(8)Separate the shortest path by means of gel electrophoresis.

(9)Sequence the shortest path to determinate its nucleotide sequence, which corresponds to the optimal solution.

4.4 Comparison with Previous Methods

Currently, there exists previous work to encode weights in DNA computing. Narayanan et al[4] presented a conceptual encoding method that encodes weights by the lengths of DNA strands. Shin et al[5] proposed a method of encoding weights by varying the number of hydrogen bonds in fixe-length DNA strands. Yamamuto et al[6] proposed a method of encoding weights by the concentrations of DNA strands. Lee et al[7] proposed a method of encoding weights by the melting temperatures of DNA strands. A ll the above methods are suitable for some specific instances, but the generated random paths in a single ligation reaction are alternant DNA strand and double-strand. The proposed DNA encoding method based on general line graph uses one DNA stand *sj* of length *w'j* to encode each vertex *v 'j,* and uses one DNA strand *s.ij= - h(s"i s'j)* to encode each edge *e'ij = (v'i, v'j).* This makes our method generate DNA double strands instead of aternant DNA strand and double strand in a single ligation reaction. Therefore, the proposed DNA encoding method can generate the optimal solution more easily than the formers, which improves the capability of representing and dealing with data in DNA computing.

5. DNA Computing Method Based on Reverse Complement Alignment for MST

For a connected and undirected graph *G = (V, E*) with *n* vertices, a spanning tree is a subgraph that contains *n* vertices and n- 1 edges with-out circle. For a weighted, connected and undireced graph, a minimum spanning tree (M ST) is a spanning tree with the minimum weight. The MST problem is to find a minimum spanning tree of the given graph. This section gives a DNA encoding method and DNA algorithm based on reverse complement alignment for the MST problem.

5.1 Definitions of Complement Alignment and Re-verse Complement Alignment

Sequence alignment reveals the relations between the characters in different sequences. The alignment a between sequences *s* and *t* is the sequence pair(*s'* , *t'*) which satisfies that[8]: (1) $|s'|=|t'|$, where $|x|$ denotes the length of sequence x;(2)after deleting all the spaces in *s* , the remaning part is *s*; (3) after deleting all the spaces in t', the remaining part is *t;* (4) for each *i, s' [i]* or t' *[i] i*s not a space. For any sequence alignment, the relation between two characters is same or diferent; whereas in DNA do ublestrand, the relation between two characters is complement, same, or different. So complement alignment and reverse complement alignment of DNA sequences are fitter than sequence alignment in DNA computing.

Definition 4. Given an alignment *a= (s', t')* between DNA sequences *s* and *t*. The alignment a is called a complement alignment[16] if a satisfies:(1)For each *i, s '[i]* is complementary to *t' [i]* if *s' [i]* and *t' [i]* are not spaces; (2) The orientation of *s'* is the same as that of *t '*.Definition 5. Given an alignment *a= (s', t')* between DNA sequences *s* and *t*. The alignment *a* is called a reverse complement alignment[16] if a satisfies: (2) For each *i, s '[i]* is complementary to *t '[i]* if *s' [i]* and *t '[i]* are not spaces; (2) The orientation of *s* is opposite to that of *t'* .

The method of computing the score of a sequence alignment is as follows[8] : For any alignment between two sequences, a bit is assigned to+1 if the two corresponding characters are the same ones (a match) ; a bit is assigned to - 1 if the two corresponding characters are different (a mismatch) ; a bit is assigned to-2 if one of the two corresponding characters is a space. T he sum of the values of all bits in an alignment is called the score of the alignment. The alignment with the biggest score is called the best alignment.

Based on the method of computing the score of sequence alignment, the method of computing the

scores of complement alignment and reverse complement alignment is given as follows. For a complement alignment or a reverse complement alignment, a bit is assigned to + 1 if the two corrsponding characters are complementary (a complment match or a reverse complement match) ; a bit is assigned to - 1 if the two corresponding chara ters are not complementary (a complement mismatch or a reverse complement mismatch) ; a bit is assigned to - 2 if one of the two corresponding characters is a space. The sum of the values of all bits in a complement alignment or reverse complement alignment is called the score of the complement alignment or reverse complement alignment. The complement alignment with the biggest score is called the best complement alignment. The reverse complement alignment with the biggest score is called the best reverse complement alignment.

5.2 DNA Encoding Method Based on Reverse Complement Alignment for MST

Definition 6. For any graph *G = (V,* E) with *n* vertices, the DNA strand used to distinguish a vertex from other *n*- 1 vertices is called the recognition code of the vertex.

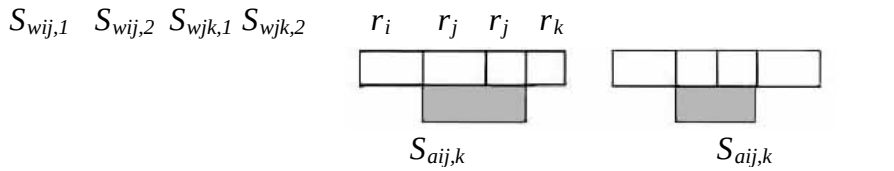
For a weighted, connected and undirected graph *G = (V, E),* v i∈*V, eij* ∈E, 1≤i,j≤n, where the weight on eij is w ij, wij≧0, wij∈Z, the length l of recognition code should satisfy: $4^{l-1} < n$ F 4^{l}, that is, $l = \lceil \log_4 n \rceil$, where 4 is the number of lelters in {*A, T, G, C*}. The DNA encoding method based on reverse complement alignment for the MST problem is given as follows:

（1）Let $l = \max\{\lceil \log_4 n \rceil, 6\}$. For each vertex

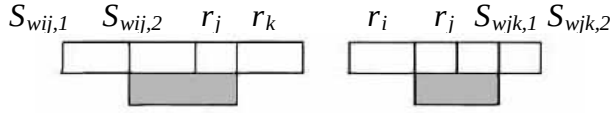
in *G,* use DNA strand *ri* of length *l* to encode it. Here, 6 is an empirical value that indicates the minimum length of recognition code.

（2）For each edge *eij* in *G,* use DNA strand *sij* of length *2p* = 2 * max{ *w ij,* l} to encode it. Here, the first part of length *l* is the same as *n,* and the last part of length *l* is the same as *rj*. And then compute the reverse complement alignment. Tf *wij* > l, then let $s_{w_{ij,1}}$ denote the first part of length *wij* of *sij,* $s_{w_{ij,2}}$ denote the last part of length *wij* of *sij*, and compute the reverse complement alignments of $s_{w_{ij,1}}$, and $s_{w_{ij,2}}$, which are marked with $\alpha_{s_{w_{ij,1}}}$ and $\alpha_{s_{w_{ij,2}}}$ respectively; otherwise, compute the reverse complement alignments of *ri* and *rj* , which are marked with α_{r_i} and α_{r_j}, respectively. Note that, when *wij* is larger than l, *ri* or *rj* cannot be the substring of the center part of *sij* .

(3)For any two adjacent edges *eij* and *ejk,* add DNA strand $sa_{ij,k}$ as an additional code. Let Upper (a) denote the upper strand of a, Lower(a) denote the lower strand of a, and " + " denote concatenation. If $wij > l$ and $wjk > l$,then the additional code $s_{a_{ij,k}} = Lower(a_{s_{wij,2}}) + Lower(a_{s_{wjk,1}})$； if wij≤l and wjk≤l , then $s_{a_{ij,k}} = Lower(a_{s_{w12,2}}) + Lower(a_{F_2})$ ； if wij>l and wjk≤l , then $s_{a_{ij,k}} = Lower(a_{s_{wij,2}}) + Lower(a_{F_j})$； if wij≤l and wjk>l， then $s_{a_{ij,k}} = Lower(a_{F_j}) + Lower(a_{s_{wjk,1}})$ Thus, DNA strands *sij* and *sjk* can combine with $s_{a_{ij,k}}$ to form DNA double strands, as shown in Fig. 3. Note that, for edges *eij* and *eji*, also add one DNA strand $s_{a_{ij,i}}$ as an additional code.



(a) $W_{ij}>1$ *and* $w_{jk}>1$ *(b)* $W_{ij}\leq 1$ *and* $w_{jk}\leq 1$



(c) $Wi_j>1$ *and* $w_{jk}\leq 1$ *(d)* $W_{ij}\leq 1$ *and* $w_{jk}>1$

Fig. 3 Joint of DNA strands S_{ij} ,S_{jk} ,and $S_{aij,k}$

Take the weighted graph *G* shown in Fig. 4 as an example. T here are 8 vertices and 12 edges in it, so $l=\max\{\lceil\log^4 8\rceil,6\}=6$ .For each vertex in *G*, use a DNA strand of length 6 to encode it. For each edge *ejij* in *G*, if *wij*≤ 6, DNA strand *sij* is with length of 12, where the first half of *sij* is ri and the last half of *sij* is *rj*. If *wij* > 6, DNA strand *sij* is with length of *2w ij*, where the first part of length 6 is the same as *ri*, the last part of length 6 is the same as *rj*, and the center part of *sij* is randomly selected but it must exclude *ri.* or *rj*. And then compute the reverse complement alignments: If *wij* ≤ 6, then compute α_{r_i} . and α_{r_j} ; otherwise,compute $\alpha_{s_{w_{ij},1}}$ and $\alpha_{s_{w_{ij},2}}$. For the edge e23, con-pute and a_3 since *w* 23 F 6; for the edge en,

compute $\alpha_{s_{w12,1}}$ and $\alpha_{s_{w12,\ 2}}$. since *w* 12>6, and so on.For any two adjacent edges *eij* and *ejk*, add one DNA strand $s_{a_{ij,k}}$ as an additional code. Consider the edges e12 and e23, the additional $c_{s_{a_{12,3}}}$ = $s_{a_{ij,k}}=Lower(a_{s_{w12,2}})+Lower(a_{F_2})$ since *w* 12 > 6 and

w 23 ≤ 6.

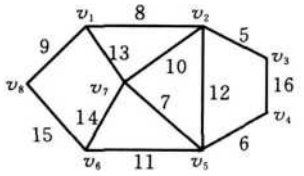


Fig. 4 A weighted an dundirected graph *G = (V, E)*

5.3 DNA Algorithm Based on Reverse Complement Alignment

For a weighted, connected and undirected graph *G = (V, E),* v i∈*V, eij* ∈*E*, 1≤*i,j*≤*n*, where the weight on edge *eij* is *wij*., the DNA algorithm based on reverse complement alignment for the M ST problem is given as follows:

(1) Let $l=\max\{\lceil\log_4 n\rceil,6\}$. For each vertex in G, use DNA strand ri of length l to encode it.

(2) For each edge eij in G, use DNA strands sij of length 2p = 2 *max { wij, I} to encode it, where the first part of length l is the same as ri, and the last part of length l is the same as rj. If wij < l, then compute the reverse complement alignments α_{r_i} and α_{r_j} . ; otherwise, compute the reverse complement alignments $\alpha_{s_{w_{ij},1}}$ and as $\alpha_{s_{w_{ij},2}}$.

(3) For any two adjacent edges *eij* and *ejk*, use DNA strand $sa_{ij,k}$ = $s_{a_{ij,k}}=Lower(a_{r_{w_{ij},2}})+Lower(a_{r_j})$ as an additional code, where ‘ *a* | b’. denotes ether *a* or *b*. If $w_{ij} > I$, then select $\alpha_{s_{w_{jk},1}}$, otherwise [w]y, 2 select α_{r_j} . ; if *wjk* > *I*, then select , otherwise select α_{r_j}

(4) Merge: DNA strands *sij* and $s_{a_{ij,k}}$ are mixed together in a single ligation reaction. Based on the principle of Watsoi-Crick complementary bases, randomly form various DNA double strands.

(5) Denature: DNA double strands are changed into DNA strands by heating.

(6) - (*N*, $s^a_{ij,k}$) : discard all the DN A strands with additional codes. Note that, all the upper strands

in the DNA encoding method do exclude additional codes.

(7) + (*N*, ri): For each vertex *v i* ($1 \le i \le n$), keep all the DNA strands with recognition code *ri*.

(8) Separate the DNA strands with the minmum weight through gel electrophoresis.

(9) Determine the nucleotide sequence of the DNA strand with the minimum weight. The sequence corresponds to an Euler cycle of the graph, and a M ST can be obtained from the Euler cycle.

6 DNA Computing Method Based on Polynomial Transformation for VCP

For an undirected graph G = (V, E) and a positive integer p, a vertex subset V A V is called a vertex cover if each edge in E is incident to at least one vert ex in V'. *The* vertex cover problem (VCP) is to find a vertex cover V' of size p in an undireced graph G. T his section presents a hybrid-based DNA solution to the vertex cover problem through designing an improved polynomial transformation from the vertex cover problem to the Hamiltonian circle problem (H CP) .

6.1 Concepts of Improved Cover Subgraph and Selection Vertex

Definition 7. For any edge e = (v_i, v_k) in an undirected graph G = (V, E), a newly constructed subgraph shown in Fig 5 is called the improved cover subgraph of e_i[17].

$(V_i,e_j,1)(V_i,e_j,2)$

$(V_k,e_j,1)(V_k,e_j,2)$

Fig. 5 T he improved cover subgraph of edge *ej = (v_i, v_k)*

The improved cover subgraph of edge *e = (vi, vk)* has 4 vertices and 4 edges. Each vert ex in it is marked with(*vi, ej,x*) or (*vk, ej,x*), $1 \le x \le 2$. The set of vertices in the improved cover subgraph of *ej* is marked with *Ve. = {(vi, ej,* 1*), (vi, ej, 2), (vk, ej,* 1*) , (v k, ej* , 2) } , and the set of edges in it is marked with *Ee. = { ((v ，ej ，*1*), (v>, e ，*2*) 4, 3(vk, e,* 1*), (vk, e,* 2*)4, 3(v,, e,* 1*), (vk, e,* 1 *)4, 3(v,, e,* 2*), (vk, e,* 2)〉 }.

Definition 8. For an undirected graph *G = (V, E)* and a positive integer *p*, the vertices in a subset *V* of size *p* are called the selection vertices[18] if the subset *V* is a vertex cover.

6. 2 Polynomial Transformation fromVCP to HCP

For an undirected graph *G = (V, E),* v i∈*V, eij* ∈*E,* 1≤i≤n,1≤j≤m and a positive integer , the polynomial transformation from the vertex cover problem to the Hamiltonian circle problem is devised as follows.

（1）Construct the improved cover subgraphs: For each edge ej = (vi, v k) , the improved cover subgraph of e is constructed, as shown in Fig. 5.

（2）Link the improved cover subgraphs of edges in G to form $|V|$ subpaths: Let deg(vi) denote the degree of vertex vi, and $e_{v_i}[1], e_{v_i}[2], \ldots, e_{v_i}[\deg(v_i)]$ denote the edges incident to v. The improved cover subgraphs of $e_{v_i}[1], e_{v_i}[2], \ldots, e_{v_i}[\deg(v_i)]$ are linked by edge set $E_{v_i} = \{\langle (vi, e_{v_i}[j], 2), (vi, e_{v_i}[j+1], 1) \rangle | 1 \le j < \deg(vi)\}$ to form one sub pat h, as shown in Fig. 6. For any undirected graph *G = (V, E)* , the number of subpaths is | *V*| , which in-clude all the improved cover subgraphs.

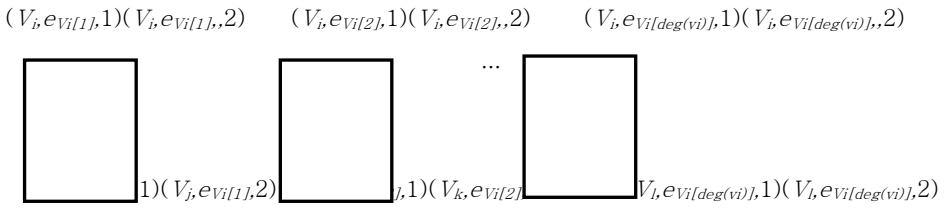


Fig. 6 The connections between the improved cover subgraphs of the edges incident to v_i

（3）Link p selection vertices to all the subpaths: Let a1, a_2,…,a_p denote p selection vertices. Link each selection vertex a_j to the start point(v_i,e_{vi}[1],1) and the end point (v_i,e_{vi}[deg(v_i)],2) of each subpath by edge set Ea_j = { <(a_j, (v_i ,e_{vi}[1],1) >, <a_j, (v_i, e_{vi}[deg(v_i)],2)〉 | 1≤ j ≤p, 1 ≤i ≤ n}. The obtained graph is marked with G =(V' , E') , where

For an undirected graph *G = (V, E*) with *n* vertices and an integer *p*≤*n*, the time complexity of the above transformation is analyzed as follows:

So the transformation can be completed in a polynomial time.

Take the graph *G = (V, E*) shown in Fig. 7 and the positive integer *p* = 2 as an example. The improved cover subgraphs of e1, *ei, e3, e4, e5* are first constructed, and then they are linked by edge set $E_{v_i}=\left\{\left\langle(vi,e_{v_i}[j],2),(vi,e_{v_i}[j+1],1)\right\rangle \mid 1\le j<\deg(vi)\right\}$ to form 4 subpaths. For vertex v 1, the improved cover subgraphs of e1, e3, *e4* are linked together to form one subpath by edges ((*v* 1, *e1*, 2*), (v* 1, e3, 1)〉 and *((v* 1, e3, *2), (v* 1, e4, 1)〉 ； for vertex *v*i, the improved cover subgraphs of e1 and ei are linked together to form one subpath by edge ((v^2, e1, 2*), (v*i, *e*i, 1)〉 , and so on. And then, all the subpaths are linked with selection vertices by edges $E_{a_j}=\left\{\left\langle a_j,\left(vi,e_{v_i}[1],1\right)\right\rangle,\left\langle aj,\left(vi,e_{v_i},e_{v_i}\left[\deg(v_i)\right],2\right)\right\rangle \mid 1\le j\le p,1\le i\le n\right\}$, that is, { (a1, *(* v 1, e1, 1*)*〉, *(a* 1, *(v* 1, e4, *2)*〉, *(a* 1, *(* v i, e1, 1)〉 , *(a* 1, *(v* i, ei, 2*)*〉,…, *(ai, (v*4, *e*4, 1*)*〉, *(a*i, *(v*4, e5, 1)〉 }. The obtained graph *G = (V , E*) is shown in Fig. 8.

$$|E|\le n(n-1)/2=0(n^2),|V'|=p+4|E|\le p+4n(n-1)/2=0(n^2),|E'|=4|E|+$$

$$\sum_{v_i\in V}\left(\deg(vi)-1\right)+2p|V|=4|E|+(2|E|-|V|)+2p|V|=6|E|+(2p-1)|V|=0(n^2)$$

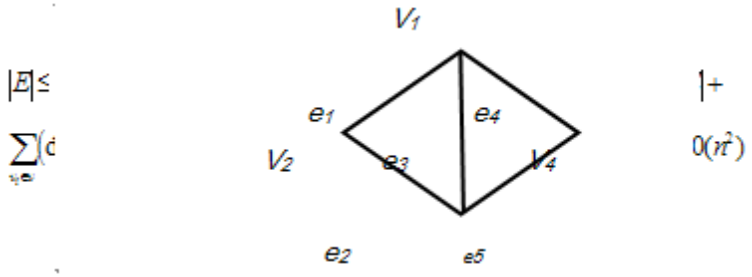


Fig. 7 A weighted and undirected graph *G = (V, E)*

It can be proved that for any undirected graph *G= (V, E*) and a positive integer *p*, the graph *G* has a vertex cover *V* of size *p* if and only if *G* =(*V* , *E*) obtained from the improved polynomial transformation has a Hamiltonian circle, where. $V'=\left\{\bigcup_{e_i\in E}V_{e_i}\right\}\cup\left\{\bigcup_{j=1\sim p}\{aj\}\right\},E'=\left\{\bigcup_{e_i\in E}E_{e_i}\right\}\cup\left\{\bigcup_{v_i\in E}E_{v_i}\right\}\cup\left\{\bigcup_{j=1\sim p}E_{aj}\right\}$

For thedetails, please see the reference [18].

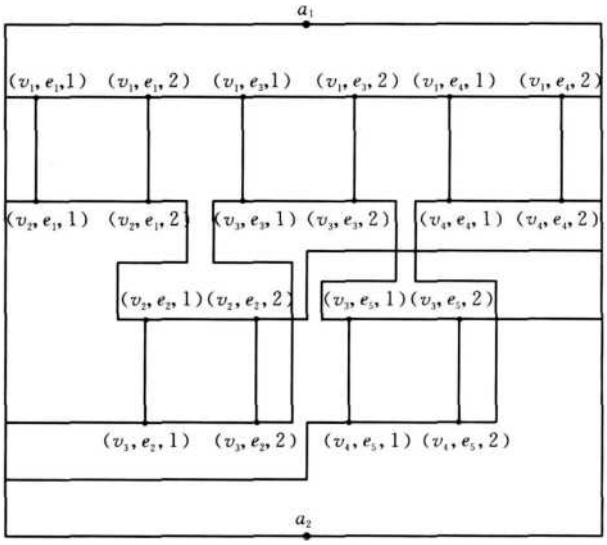


Fig. 8 T he connect ions betw een the improved cover subgraphs of the edges in graph G

6.3 DNA Encoding Method Based on Polynomial Transformation

For an undirected graph G = (V, E) and a positive integer p , the DNA encoding method based on the improved polynomial transformation for the vertex cover problem is given as follows:

(1) For each edge ej = (vi, vk) in G, construct the improved cover subgraph of ej. For all the edges incident to Vi, the improved cover subgraphs are linked together to form one subpath by edge set $E_{v_i} = \left\{ \left\langle (vi, e_{v_i}[j], 2), (vi, e_{v_i}[j+1], 1) \right\rangle \mid 1 \le j < \deg(vi) \right\}$. The number of the subpaths is | V | , which include all the improved cover subgraphs of the edges in G. Link p selection vertices to the start and end points of each subpath by edge set $E_{a_j} = \left\{ \left\langle a_j, (vi, e_{v_i}[1], 1) \right\rangle, \left\langle aj, \begin{pmatrix} vi, e_{v_i}, e_{v_i}[\deg(v_i)], \\ 2 \end{pmatrix} \right\rangle \mid 1 \le j \le p, 1 \le i \le n \right\}$ .The obtained graph is marked with G' = (V,' E'), where $v' = \left\{ \bigcup_{e_i \in E} V_{e_i} \right\} \bigcup \left\{ \bigcup_{j=1\sim p} \{aj\} \right\}, E' = \left\{ \bigcup_{e_i \in E} E_{e_i} \right\} \bigcup \left\{ \bigcup_{v_i \in E} E_{v_i} \right\} \bigcup \left\{ \bigcup_{j=1\sim p} E_{aj} \right\}$

(2) For each vertex v_i in G' , use DNA stand s_i of length l to encode it. Here, select l = 20.

(3) For each edge $e^j = (v_i, v_k)$ in G' , use the DNA strands $s_{ik} = - h(s''_i s'_j)$ to encode it.

6. 4 DNA Algorithm Based on Polynomial Tran- formation

For an undirected graph *G = (V, E)* and a pos-itive integer *p* , the DNA algorithm based on the polynomial transformation for the vertex cover problem is given as follows.

（1） Convert G = (V, E) to G' = (V', E') through the improved polynomial transformation, where $v' = \left\{ \bigcup_{e_i \in E} V_{e_i} \right\} \bigcup \left\{ \bigcup_{j=1\sim p} \{aj\} \right\}, E' = \left\{ \bigcup_{e_i \in E} E_{e_i} \right\} \bigcup \left\{ \bigcup_{v_i \in E} E_{v_i} \right\} \bigcup \left\{ \bigcup_{j=1\sim p} E_{aj} \right\}$. For each vertex v_i in G' ,use DNA strand s_i of length l = 20 to encode it. For each edge e_j = (v_i, v_k) in G' , use the DNA strand s_{ik} = - *h(s"$_i$ s'$_k$)* to encode it.

(2)Merge: DNA strands si and sij are mixed together in a single ligation reaction. Based on the principle of Watson-Crick complementary bases, randomly form various DNA double strands.

(3)Denature: DNA double strands are changed into DNA strands by heating.

（4）(N≤ 20 × n'): keep all the DNA molecules whose length are less than or equal to 20× n', where n' is the number of vertices in G'.

(5) + (N, s_i) : For each vertex vi in G' , keep all the DNA molecules with s_i.

(6) Detect: If any DNA molecule remains, say "Yes"； otherwise say " No" . If the answer is "Yes", then separate the DNA molecules obtained in step 5 through agarose gel electrophoresis, and the furthest DNA molecules are just what we want. Sequencing the furthest DNA molecule to determinate its nucleotide sequence, which corresponds to the optimal solution.

6.5 Comparison with Previous Methods

For any instance of the vertex cover problem, the previous DNA algorithms[19-20] are based on sticker model, and the proposed DNA algorithm is based on hybrid system. Firstly, a polynomial transformation is devised to convert any instance of the vertex cover problem to that of the Hamilton-

an circle problem. And then give a DNA encoding method and DNA algorithm based on hybridization for the vertex cover problem. The operations used in the method are implemented more easily than that in the previous methods. Compared with the previous polynomial transformation[18], the inproved cover subgraph has 4 vertices and 4 edges instead of 12 vertices and 14 edges. Thus, the number of vertices in the obtained graph *G'* is reduced to $p + 4|E|$from $p + 12|E|$, and the number of edges is reduced to $6|E| + (2p - 1)|V|$from $16|E|+ (2p - 1)|V|$.

7. Conclusion

This paper discusses DNA computing methods of optimization problems on weighted graph. The proposed DNA encoding method based on relative length graph for the traveling salesman problem can directly deal with weights of real numbers and integers, and can distinguish the paths with almost equal weights. The obtained solution is not proportional to the length of DNA strand, so it can encode weights of broad range. The designed DNA encoding method based on general line graph for the Chinese postman problem uses two DNA stands of different lengths to encode each vertex. T his makes the codes of weights on edges are dealt with by the method similar to the codes of vertces, so the weights on edges are dealt with more easily than in the formers. The presented DNA encoding method based on reverse complement alignment for the minimum spanning tree problem selects the additional codes by computing the reverse complement alignment of the first and last parts of edge codes, and finds the minimum spanning tree by getting an Euler circle of the given instance. The proposed DNA encoding method based on poynomial transformation for the vertex cover problem is an improvement on the previous method. T he number of vertices and edges in the improved cover subgraph is decreased, and so the space complexity is reduced.

References

[1]Adleman L M. Molecular computation of solutions to combnatorial problems. Science, 1994, 266: 1021— 1024

[2] Lipton R J. DNA solution of hard computational problems. Science, 1995, 268: 542— 545

[3]Ouyang Q, Kaplan P D, Liu S et al. DNA solution of the maximal clique problem . Science, 1997, 278: 446— 449

[4]Narayanan A, Zorbalas S et al. DNA algorithm s for computing shortest paths//Proceedings of the 3rd Annual Genetic Programming Conference. San Francisco: Morgan Kaufma in Publisher, 1998: 718— 723

[5]Shin S Y, Zhang B T , Jun S S et al. Solving traveling saleman problem s using molecular program ming//Proceedings of the 1999 Congress on Evolutionary Computation. Piscaaway: IEEE Press, 1999: 994- 1000

[6]Yam am oto M, H iroto Y , Matoba T . Solutions of shortest path problems by concentration control//Lecture Notes in Computer Science 2340, Berlin: Springer, 2002: 231-240

[7]Lee J Y, Shin S Y, Park T H et al. Solving traveling salesman problem s w it h DNA molecules en co ding numerical vaues. BioSystems, 2004, 78: 39-47

[8] Setubal J , MeidanisJ. Introduction to Computational Moleular Biology. Thomson: Cole Publishing Company, 1997 (etubal J, Meidanis J.)

[9] Plun G, Rozenberg G, Salomaa A. DN A Computing: New Computing Paradigms. Berlin: Springer, 1998 Plun G, Rozenberg G, Salomaa A.

[10] Xu J in, Zhang She- Min, Fan Y ue- Ke, Guo Yang— An. DNA com puter principle, advances and difficulties（Ⅲ）：Thestructure and character of data in DNA computing.Chinese Journal of

Computers, 2007,30(6);869-880(ih Chinese)

[11]Han Ai-Li, Zhu Da-Ming. DNA Encoding Mcthals in the Field of DNA (Computing// Computational Tntdligcncc in Bioin forma tics, Springcr- Vcrlag, 2008

[12]Han Ai-Li, Zhu Da- Ming. A now DNA- based approach to soive the maximum weight clique problem //Lceture Notes in computer Science 4115.Bcrlin:Spring,2006:320-327

[13]Han Ai-Li. RLM: A new method of encoding wei ghts in DNA strands// Proceedings of the International conference on Hybrid Tntelligent Systems. Piscatawa IEEE Press,2006: 122- 125

[14]Han Ai-Li, Zhu Da-Ming. A new DNA encoding method for [19] travel ing salesman problem/ / Lecture Notes in Computer Scence 4115. Berlin: Springer, 2006: 328-335

[15]Han Ai-Li, Z hu D a- M ing. DN A computing model based on a new schem *e* of encoding weight for Chinese postman problem. Journal of Computer Rcscarch and Development, 2007,

44(6) : 1053 ; - 1062(in Chinese)

[16]Han Ai-Li, Zhu Da-Ming. DNA solution based on sequence alignment to the minimum spanning tree problem. International Journal of Bioinformatics Rescarch and Applications,2008

4(2): 188-200

[17]Han Ai-Li, Zhu Da-M ing. Hybrid-based DNA solution to the vertex cover problem//Proeecdings of the 2nd International Symposium on Tntdligence Computation and Applications. Wuhan, China, 2007: 87-91

[18]Ma Shao-Han. Analysis and Design of Algorithms. Jinan: Shandong University Press, 1992(in Chinese)

[19]Gao Lin, M a Run-Nian, Xu Jin. DNA solutions of vertex cover problem based on sticker model. Chinese Electronics,

2002, 11(2): 280-284

[20]Chen Zhi-Yun, Qu Hui-Qin, Lu MingrMing, Zhu Hong. A probabilistic parameterized algorithm for vertex cover in sticker sodel//Proceedings of the 18th International Parallel and Distributed Prossing Symposium. 2004, 18 : 2359-2363