Population Diversity Measures Based on Variable-Order Markov Models for the Traveling Salesman Problem

Yuichi Nagata(✉)

Graduate School of Science and Technology, Tokushima University, Tokushima, Japan
nagata@is.tokushima-u.ac.jp

Abstract. This paper presents entropy-based population diversity measures that take into account dependencies between the variables in order to maintain genetic diversity in a GA for the traveling salesman problem. The first one is formulated as the entropy rate of a variable-order Markov process, where the probability of occurrence of each vertex is assumed to be dependent on the preceding vertices of variable length in the population. Compared to the use of a fixed-order Markov model, the variable-order model has the advantage of avoiding the lack of sufficient statistics for the estimation of the exponentially increasing number of conditional probability components as the order of the Markov process increases. Moreover, we develop a more elaborate population diversity measure by further reducing the problem of the lack of statistics

1 Introduction

Maintaining the genetic diversity in the population is one of the most important factors for bringing out the potential of genetic algorithms (GAs). One of the approaches to maintain population diversity is to design an appropriate measure of population diversity, which are used as a trigger to activate diversification procedures [8,9] and a part of the fitness function to maintain population diversity in a positive manner [3,5,11].

As is well known in information theory, entropy is a measure of the uncertainty of a probability distribution and it has been used to design population diversity measures. Most of the entropy-based population diversity measures are defined as the sum of the entropies of the univariate marginal distributions of all variables in the form of $-\sum_{i=1}^{n}\sum_{a\in A} P(X_i = a)\log P(X_i = a)$. This type of population diversity measure is widely used in GAs for the knapsack problem [4], binary quadratic programming problem [10], traveling salesman problem [3,5,7,8], and others [11]. This entropy measure, however, does not have an ability to capture dependencies between the variables.

In our previous work [6], we proposed an entropy-based diversity measure that takes into account dependencies between the variables, and this measure was used to maintain population diversity in a GA for the traveling salesman

J. Handl et al. (Eds.): PPSN XIV 2016, LNCS 9921, pp. 973–983, 2016.
DOI: 10.1007/978-3-319-45823-6_91

problem (TSP). This diversity measure, denoted as H_m, was formulated as the entropy rate of a Markov process of order m, where the probability of occurrence of each vertex at a certain position was assumed to be dependent on the m preceding vertices in the population (tours). The use of the diversity measure H_m with an appropriate value of m $(= 4)$ improved the performance of the GA.

In the practical use of a fixed-order Markov model, there is seldom sufficient data to accurately estimate the exponentially increasing number of conditional probability components as the order of the Markov model increases. A variable-order Markov model is useful to reduce this problem, where the probability of occurrence of each symbol is assumed to be dependent on the preceding symbols of variable length, which varies depending on the available statistics. Variable-order Markov models have been successfully applied to areas such as machine learning [1] and bioinformatics [2]. In this paper, we develop an population diversity measure based on a variable-order Markov model, which models the probability distribution of individuals in the population. Moreover, we improve this diversity measure by further reducing the problem of the lack of data.

The remainder of this paper is organized as follows. In Sect. 2, we first describe the diversity measure H_m and its variant proposed in the previous work. Then, we propose two entropy-based diversity measures derived from variable-order Markov models. The GA framework used to evaluate the proposed diversity measures is described in Sect. 4. Computational results are presented in Sect. 5 and conclusion is given in Sect. 6.

2 Previous Work

In [6], we proposed an entropy-based population diversity measure that takes into account dependencies in sequences of vertices included in the population of the GA for the TSP. This section outlines this work.

Let S_i $(i = 1, \ldots, n)$ be a random variable representing the i-th vertex in the tours of the population, where n is the number of the vertices (cities). The probability of occurrence of each vertex at a certain position is modeled as a Markov process of order m, where it is assumed to be dependent on the m preceding vertices in the tours of the population. Given that each tour has a cyclic structure, the joint probability distribution $P(S_1 = s_1, S_2 = s_2, \ldots, S_n = s_n)$, which is denoted as $P(s_1, s_2, \ldots, s_n)$ for simplicity, is represented by the following formula, where index $i + n$ $(1 \le i \le m)$ corresponds to i.

$$P(s_1, s_2, \ldots, s_n) = \prod_{i=1}^{n} P(s_{i+m} \mid s_i, \ldots, s_{i+m-1}) \tag{1}$$

Given that each tour can start from an arbitrary vertex, the joint probability distribution of any subset of the sequence of random variables should be invariant with respect to shifts in the index. Therefore, the entropy H of this joint probability distribution is equivalent to nH_m (Eq. 2), where H_m (Eq. 3) is the entropy rate of the Markov process of order m that models the probability of occurrence of each vertex in the population. For a more detailed explanation of Eq. 2, see the previous work. Equation 3 can be easily transformed into Eq. 5.

In information theory, the entropy rate of a data source is the average number of bits per symbol needed to encode it. Therefore, the existence of the same sequence consisting of up to $m+1$ vertices in the population will decrease the value of H_m.

$$H = -\sum_{s_1} \cdots \sum_{s_n} P(s_1, \ldots, s_n) \log P(s_1, \ldots, s_n) = nH_m \tag{2}$$

$$\begin{aligned} H_m &= -\sum_{s_1} \cdots \sum_{s_{m+1}} P(s_1, \ldots, s_{m+1}) \log P(s_{m+1} \mid s_1, \ldots, s_m) && (3)\\ &= -\sum_{s_1} \cdots \sum_{s_{m+1}} P(s_1, \ldots, s_{m+1}) \log \frac{P(s_1, \ldots, s_{m+1})}{P(s_1, \ldots, s_m)} && (4)\\ &= \overline{H_{m+1}} - \overline{H_m}, && (5) \end{aligned}$$

where

$$\overline{H_k} = -\sum_{s_1} \cdots \sum_{s_k} P(s_1, \ldots, s_k) \log P(s_1, \ldots, s_k). \tag{6}$$

To compute $\overline{H_k}$ in the asymmetric TSP, all sequences of length k are sampled in the population, and $P(s_1, \ldots, s_k)$ is estimated by $\frac{N(s_1, \ldots, s_k)}{nN_{pop}}$, where $N(s_1, \ldots, s_k)$ is the number of a sequence of vertices $\{s_1, \ldots, s_k\}$ in the population consisting of N_{pop} tours. In the symmetric TSP, the sampling is conducted in both travel directions and $P(s_1, \ldots, s_k)$ is estimated by $\frac{N(s_1, \ldots, s_k)}{2nN_{pop}}$.

Another diversity measure, denoted as H'_m, was also proposed. This measure is defined as the sum of the diversity measures H_k $(k = 1, \ldots, m)$, which can be simplified as Eq. 7. This diversity measure was designed in an ad hoc way to reduce the problem of the lack of statistics for the accurate estimation of H_m.

$$H'_m = H_1 + H_2 + \cdots + H_m = \overline{H_{m+1}} - \overline{H_1} \tag{7}$$

3 Population Diversity Measures Based on Variable-Order Markov Models

3.1 Motivation

The population diversity measure proposed in this paper is also defined as the entropy rate of a Markov process. We denote a set of the symbols generated from an information source as L. In what follows, we use random variables S_i $(i = \ldots, -2, -1, 0)$ to represent a Markov process, where S_0 represents the symbol to be observed next and S_{-i} $(i > 0)$ represents the i-th preceding symbol. The expression of H_m is therefore given by the following formula.

$$H_m = -\sum_{s_{-m}} \cdots \sum_{s_{-1}} \sum_{s_0} P(s_{-m}, \ldots, s_{-1}, s_0) \log P(s_0 \mid s_{-m}, \ldots, s_{-1}) \tag{8}$$

In theory, the value of H_m gives the entropy rate of an Markov process of order k (*i.e.*, $H_m = H_k$) as long as $k \leq m$ because $P(s_0 \mid s_{-m}, \ldots, s_{-k}, \ldots, s_{-1}) = P(s_0 \mid s_{-k}, \ldots, s_{-1})$ in this case. Therefore, m should be set to a greater value so that the entropy rate H_m has an ability to capture higher-order dependencies in a sequence of symbols generated from an information source. If m is too large, however, H_m would not be a meaningful population diversity measure because there is seldom sufficient samples of sequences in the population to accurately estimate the conditional probability distributions $P(s_0 \mid s_{-m}, \ldots, s_{-1})$, $s_{-m}, \ldots, s_0 \in L$, which are estimated as $\frac{N(s_{-m},\ldots,s_{-1},s_0)}{N(s_{-m},\ldots,s_{-1})}$. Therefore, there is a tradeoff between the potential ability to capture higher-order dependencies and the estimate accuracy of the conditional probability distributions. The population diversity measures proposed in this paper aim to capture higher-order dependencies in sequences of vertices in the population while reducing the problem of the lack of data.

3.2 A Population Diversity Measure H_m^{tr1}

We model the probability of occurrence of a symbol (vertex) appearing in sequences of symbols (sequences of vertices in the population) as a variable-order Markov process. In a variable-order Markov process, the probability distribution of the next symbol s_0 depends on the preceding symbols of variable length k. The basic idea is to determine the value of k adaptively so that the number of samples $N(s_{-k}, ..., s_{-1})$ is a sufficient statistic for estimating the conditional probability distribution $P(s_0 \mid s_{-k}, \ldots, s_{-1})$. For example, if a specific sequence of symbols $\{\ldots, s'_{-3}, s'_{-2}, s'_{-1}\}$ is observed at a certain point, the conditional probability distribution of occurrence of the next symbol s_0 is modeled as $P(s_0|s'_{-k}, ..., s'_{-1})$ such that the number of samples $N(s'_{-k}, ..., s'_{-1})$ is greater than a predefined minimum number of samples.

A variable-order Markov process is characterized by a set of the conditional probability distributions: $P(s_0|\boldsymbol{s}_c)$, $\boldsymbol{s}_c \in \mathcal{S}$, where $\mathcal{S}$ is a set of sequences of symbols for the conditioning variables and each element $\boldsymbol{s}_c$ represents a specific sequence of symbols of any length that is less than or equal to m. Here, we put the upper limit on the length of sequences for the conditioning variables because it is impractical to store all conditional probability components if m is too large (e.g. $m > 10$). For any sequence of symbols $\{\ldots, s_{-2}, s_{-1}\}$ at a certain point, the length of the sequence assigned to the conditioning variables must be uniquely determined. To represent set $\mathcal{S}$ that satisfies this requirement, a so-called *context tree* is useful. Let $\tilde{\boldsymbol{s}_c}$ be the reverse sequence of $\boldsymbol{s}_c$ and $\tilde{\mathcal{S}} = \{\tilde{\boldsymbol{s}_c}|\boldsymbol{s}_c \in \mathcal{S}\}$. The elements of $\tilde{\mathcal{S}}$ are represented as the leaf nodes of a context tree as illustrated in Fig. 1 (Left), where every node has either 0 or $|L|$ children.

The entropy rate of the variable-order Markov process, which we denote as H_m^{tr1}, is then defined by the following formula.

$$H_m^{tr1} = -\sum_{\boldsymbol{s}_c \in \mathcal{S}} \sum_{s_0 \in L} P(\boldsymbol{s}_c, s_0) \log P(s_0 \mid \boldsymbol{s}_c) = -\sum_{\boldsymbol{s}_c \in \mathcal{S}} \sum_{s_0 \in L} P(\boldsymbol{s}_c, s_0) \log \frac{P(\boldsymbol{s}_c, s_0)}{P(\boldsymbol{s}_c)}$$
$$= -\sum_{\boldsymbol{s}_c \in \mathcal{S}} \sum_{s_0 \in L} P(\boldsymbol{s}_c, s_0) \log P(\boldsymbol{s}_c, s_0) + \sum_{\boldsymbol{s}_c \in \mathcal{S}} P(\boldsymbol{s}_c) \log P(\boldsymbol{s}_c) \quad (9)$$

The entropy rate H_m^{tr1} is closely related to H_m. If a context tree $\tilde{\mathcal{S}}$ is represented as a perfect tree with depth m, H_m^{tr1} is equivalent to H_m, meaning that H_m^{tr1} is a generalization of H_m. In addition, H_m^{tr1} can be viewed as an approximation of H_m. In fact, H_m^{tr1} is obtained from H_m though the approximation of $P(s_0 \mid s_{-m}, \ldots, s_{-k}, \ldots, s_{-1}) = P(s_0 \mid s_{-k}, \ldots, s_{-1})$ for all $\{s_{-k}, \ldots, s_{-1}\} \in \mathcal{S}$.

Next, we describe how to determine set $\tilde{\mathcal{S}}$ (and equivalently $\mathcal{S}$). The corresponding context tree $\tilde{\mathcal{S}}$ is updated at fixed intervals (see Sect. 4) by the following procedure, where *ratio* is a parameter taking a value between 0 and 1.

1. $\tilde{\mathcal{S}}$ is initialized as the perfect tree of depth one, *i.e.*, $\tilde{\mathcal{S}} = \{s_{-1} | s_{-1} \in L\}$.
2. For each of the leaf nodes $\{s_{-1}, ..., s_{-k}\} \in \tilde{\mathcal{S}}$, if there exists at least one value $s'_{-(k+1)} \in L$ such that $N(s'_{-(k+1)}, s_{-k}, \ldots, s_{-1}) \geq N_{pop} * ratio$, this node is expanded to generate the new leaf nodes $\{s_{-1}, ..., s_{-k}, s_{-(k+1)}\}$, $s_{-(k+1)} \in L$.
3. Expansions of the leaf nodes are iterated until no expansion is possible or the depth of each leaf node reaches the predefined maximum number m. The resulting tree $\tilde{\mathcal{S}}$ is returned.

The aim behind the expansion of a leaf node $\{s_{-1}, ..., s_{-k}\} \in \tilde{\mathcal{S}}$ is to capture higher-order dependency expressed as the conditional probability distribution $P(s_0 | s'_{-(k+1)}, s_{-k}, \ldots, s_{-1})$ only when it is judged to be reliable. The parameter *ratio* balances the tradeoff between the potential ability to capture higher-order

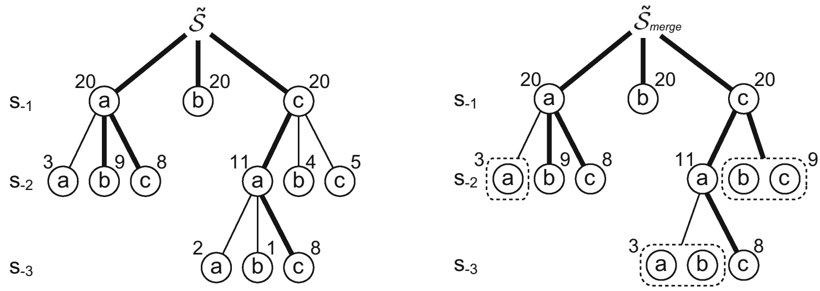

Fig. 1. (Left) A context tree representation of $\tilde{\mathcal{S}}$, where $L = \{a, b, c\}$ and the threshold is 8. Each node is connected by a thick link if the number of the corresponding sequence in the population (indicated beside each node) is greater than or equal to the threshold. The corresponding Markov process is defined as follows: $P(s_0|a, a)$, $P(s_0|b, a)$, $P(s_0|c, a)$, $P(s_0|b)$, $P(s_0|a, a, c)$, $P(s_0|b, a, c)$, $P(s_0|c, a, c)$, $P(s_0|b, c)$, and $P(s_0|c, c)$. (Right) A context tree representation of $\tilde{\mathcal{S}}_{merge}$ obtained from $\tilde{\mathcal{S}}$. Nodes in each dotted frame are merged. The corresponding Markov process is defined as follows: $P(s_0|a, a)$, $P(s_0|b, a)$, $P(s_0|c, a)$, $P(s_0|b)$, $P(s_0|a \vee b, a, c)$, $P(s_0|c, a, c)$ and $P(s_0|b \vee c, c)$.

dependencies and the estimate accuracy of the conditional probability distributions. However, this expansion collaterally generates unreliable conditional probability distributions $P(s_0|s''_{-(k+1)}, s_{-k}, \ldots, s_{-1})$ for $s''_{-(k+1)} \in L \setminus \{s'_{-(k+1)}\}$ because the values of $N(s''_{-(k+1)}, s_{-k}, \ldots, s_{-1})$ are less than the predefined threshold. Note that if expansion of a leaf node is allowed only if the number of samples is greater than the threshold for all child nodes, no expansion is likely to occur.

3.3 A Population Diversity Measure H_m^{tr2}

As suggested in the previous subsection, unreliable conditional probability distributions included in the formulation of H_m^{tr1} have a potentially harmful effect on the evaluation of population diversity. To reduce this problem, we modify the variable-order Markov model used to derive H_m^{tr1}.

The basic idea is to merge unreliable conditional probability distributions into a single one in order to increase the number of samples for the conditioning variables. One simple method is to merge the unreliable conditional parts $\{s''_{-(k+1)}, s_{-k}, \ldots, s_{-1}\}$, $s''_{-(k+1)} \in L \setminus \{s'_{-(k+1)}\}$ into a single one. Figure 1 illustrates an example where the unreliable conditional parts (nodes connected by thin links) in $\tilde{\mathcal{S}}$ are merged accordingly. We denote the resulting set of sequences of symbols for the conditioning variables and corresponding context tree as $\tilde{\mathcal{S}}_{merge}$. For example, a merged conditional probability distribution $P(s_0|b \vee c, c)$ is estimated by $\frac{N(b,c,s_0)+N(c,c,s_0)}{N(b,c)+N(c,c)}$. Although the number of samples for a merged conditional part may be still less than the predefined threshold, the problem of the lack of sufficient data will be alleviated. We denote the entropy rate of the variable-order Markov process defined by $\tilde{\mathcal{S}}_{merge}$ as H_m^{tr2}.

4 GA Framework

To evaluate the ability of the proposed population diversity measures H_m^{tr1} and H_m^{tr2}, we perform the GA proposed in [5] as in the case of the previous work [6]. Algorithm 1 gives the GA framework where brief comments are written directly in the algorithm. For more details, see the previous work [6].

An important point is that each of the population diversity measures is incorporated into the evaluation function used for selecting individuals to survive (line 8). Let L be the average tour length of the population and H the population diversity measure (H_m^{tr1} or H_m^{tr2}). For each individual $y \in \{c_1, \ldots, c_{N_{ch}}, p_A\}$, it is evaluated by the following evaluation function (Eq. 10), and the one with the smallest value is selected to replace the population member selected as p_A. Here, $\Delta L(y)$ and $\Delta H(y)$ denote the differences in L and H, respectively, when $x_{r(i)}$ $(= p_A)$ is replaced with an offspring solution y. This evaluation function is motivated to minimize $L - TH$ after the replacement, where T is a coefficient that takes a balance between the influences from L and H and it is adaptively updated (basically decreased) during the course of the search. Note that offspring

Algorithm 1. Procedure GA

Generate initial population $\{x_1, \ldots, x_{N_{pop}}\}$; // a simple local search is used;
repeat
  Update $\tilde{\mathcal{S}}$ or $\tilde{\mathcal{S}}_{merge}$ based on the procedure described in Section 3 ;
  Let $r(\cdot)$ be a random permutation of $1, \ldots, N_{pop}$;
  for $i := 1$ to N_{pop} **do**
    $p_A := x_{r(i)}$, $p_B := x_{r(i+1)}$; // set a pair of parents
    $\{c_1, \ldots, c_{N_{ch}}\} :=$ CROSSOVER(p_A, p_B); // generate N_{ch} offspring solutions using edge assembly crossover
    $x_{r(i)} :=$ SELECT_BEST$(c_1, \ldots, c_{N_{ch}}, p_A)$; // select the best individual to replaces the population member selected as p_A
  end for
until a termination condition is satisfied
return the best individual in the population;

solutions that increase L are never selected in order to prevent the population from not converging, *i.e.*, no replacement occurs when p_A itself is selected.

$$Eval(y) = \begin{cases} \Delta L(y) - T\Delta H(y) \ (\Delta L(y) \leq 0) \\ \infty \ (\Delta L(y) > 0) \end{cases} \tag{10}$$

For every offspring solution y, $\Delta H(y)$ can be computed in $O(km)$ time, where k is the number of edges of an offspring solution y that do not exist in the parent p_A (k is usually much smaller than n). Each time p_A is replaced with the selected offspring solution, the values of $N(\cdot)$, which are stored in the form of a tree, can be updated in $O(km)$ time.

5 Experimental Results

5.1 Experimental Settings

To investigate the ability of the proposed population diversity measures H_m^{tr1} and H_m^{tr2}, we performed the GA described in the previous section by using each of the population diversity measures in the evaluation function (Eq. 10). The parameters for the GA were set as follows: $N_{pop} = 300$ and $N_{ch} = 30$. Note that the same settings were used in the previous work [6] for evaluating the population diversity measures H_m and H'_m. We tested the proposed population diversity measures with the following parameter settings.

- H_m^{tr1} $(m = 6,\ ratio = 0.05,\ 0.1,\ 0.2,\ 0.3)$
- H_m^{tr2} $(m = 6,\ ratio = 0.05,\ 0.1,\ 0.2,\ 0.3)$
- H_m^{tr2} $(m = 8,\ ratio = 0.1)$

For each setting, we performed the GA 30 times on 21 instances with sizes ranging from 10,000 to 25,000 in the well-known benchmark sets: TSPLIB (http://comopt.ifi.uni-heidelberg.de/software/TSPLIB95/), National TSPs (http://www.math.uwaterloo.ca/tsp/data/index.html), and VLSI TSPs.

5.2 Results

Table 1 shows the solution quality of the GA using the proposed population diversity measures H_m^{tr1} and H_m^{tr2} in the following format: the instance name (instance) together with the optimal or best known solution (Opt. or UB), the number of runs that succeeded in finding the optimal or best-known solution (#S), and the average percentage excess over the optimal or best-known solutions (A-Err). The result of the GA using the diversity measure H_4, which achieved the best solution quality among H_m ($m = 1, 2, 3, 4, 6$) in the previous work, are also presented for a baseline comparison. We performed the one-sided Wilcoxon rank sum test for the null hypothesis that the median of the distribution of tour length obtained by the GA using each of H_m^{tr1} and H_m^{tr2} is greater than that of GA using H_4. If the null hypothesis is rejected as a significant level of 0.05, results in the table are indicated by the asterisk. In addition, results are also indicated by the dagger if the opposite null hypothesis is rejected.

Let us first summarize the results of the GA using the population diversity measures H_m and H'_m ($m = 1, 2, 3, 4, 6, 8$) proposed in the previous work. Table 2 shows only averaged results taken from [6] (results of $m = 8$ are newly added). As indicated in Table 2, the diversity measure H_m improves the ability in evaluating population diversity with increasing the value of m up to 4, but the greater values of m deteriorates the ability due to the lack of available samples necessary to estimate the conditional probability distributions. The diversity measure H'_m also improves the ability in evaluating population diversity with increasing the value of m up to 6. Moreover, the result of $H_6{}'$ is better than that of H_4. Considering the definition of H'_m, this result suggests that the diversity measure H'_6 achieves a better balance between the ability to capture higher-order dependencies and the estimate accuracy of the conditional probability distributions, although the definition of H'_m is somewhat ad-hoc.

Next, we focus on the results of the diversity measure H_6^{tr1}. Table 1 shows that the GA using H_6^{tr1} achieves the best solution quality when the parameter *ratio* is set to 0.2 or 0.3. For a smaller value of *ratio*, the solution quality is deteriorated. This is a predictable consequence because if the value of *ratio* is too small, it would not be likely to obtain a sufficient statistics from the population necessary for the accurate estimation of the conditional probability distributions. The use of H_6^{tr1} with the best parameter value for *ratio* (= 0.2 or 0.3), however, shows only a slight improvement over the use of H_4.

Next, we focus on the results of the diversity measure H_6^{tr2}. The GA using H_6^{tr2} achieves the best solution quality when the parameter *ratio* is set to 0.1, which is less than the best parameter value for H_6^{tr1}. Moreover, the best result of H_6^{tr2} is better than that of H_6^{tr1}. These observations indicate that the use of H_m^{tr2} succeeds in capturing higher-order dependencies while reducing the problem of the lack of sufficient samples for the accurate estimation of the conditional probability distributions. Compared to H_4, the use of H_6^{tr2} with the best parameter value for *ratio* (= 0.1) significantly improves the solution quality in four instances. However, this result is almost same as that of H'_6.

Table 1. Solution quality of the GA using the diversity measures H_m^{tr1} and H_m^{tr2}

		H_4		H_6^{tr1}							
				$ratio = 0.05$		$ratio = 0.1$		$ratio = 0.2$		$ratio = 0.3$	
Instance	Opt.(UB)	#S	A-Err	#S	A-Err	#S	A-Err	#S	A-Err	#S	A-Err
xmc10150	(28387)	24	0.00070	26	0.00047	29	0.00012*	29	0.00012*	28	0.00035
fi10639	520527	24	0.00011	21	0.00021	23	0.00016	23	0.00018	24	0.00010
rl11849	923288	25	0.00014	18	0.00037†	19	0.00030†	21	0.00026	27	0.00013
usa13509	19982859	22	0.00010	17	0.00017	24	0.00008	21	0.00010	19	0.00014
xvb13584	(37083)	29	0.00009	23	0.00081†	26	0.00036	27	0.00027	29	0.00009
brd14051	469385	23	0.00017	19	0.00026	26	0.00008	27	0.00005	26	0.00011
mo14185	(427377)	19	0.00014	20	0.00014	19	0.00018	18	0.00015	19	0.00016
xrb14233	(45462)	10	0.00279	9	0.00286	11	0.00279	10	0.00301	12	0.00271
d15112	1573084	16	0.00014	17	0.00008	16	0.00007	15	0.00005	17	0.00003
it16862	557315	6	0.00023	4	0.00044†	2	0.00040†	2	0.00039†	6	0.00030
xia16928	(52850)	24	0.00076	23	0.00050	18	0.00101	19	0.00095	16	0.00164†
pjh17845	(48092)	13	0.00132	17	0.00097	15	0.00125	15	0.00104	13	0.00118
d18512	645238	21	0.00009	20	0.00009	22	0.00008	23	0.00009	25	0.00007
frh19289	(55798)	30	0.00000	26	0.00030†	26	0.00024†	28	0.00012	30	0.00000
fnc19402	(59287)	19	0.00067	16	0.00079	18	0.00067	17	0.00079	19	0.00062
ido21215	(63517)	23	0.00058	18	0.00105	22	0.00058	27	0.00016	17	0.00110†
fma21553	(66527)	15	0.00090	10	0.00120	8	0.00120	16	0.00070	21	0.00050
vm22775	569288	0	0.00140	1	0.00141	0	0.00131	0	0.00121	1	0.00119
lsb22777	(60977)	21	0.00055	19	0.00060	22	0.00044	28	0.00011*	24	0.00033
xrh24104	(69294)	29	0.00005	28	0.00010	26	0.00019	28	0.00010	29	0.00005
sw24978	855597	9	0.00039	11	0.00047	14	0.00037	12	0.00031	11	0.00024
Average		19.1	0.00054	17.3	0.00063	18.4	0.00057	19.3	0.00048	19.7	0.00053

		H_6^{tr2}								H_8^{tr2}	
		$ratio = 0.05$		$ratio = 0.1$		$ratio = 0.2$		$ratio = 0.3$		$ratio = 0.1$	
Instance	Opt.(UB)	#S	A-Err	#S	A-Err	#S	A-Err	#S	A-Err	#S	A-Err
xmc10150	(28387)	26	0.00059	25	0.00070	28	0.00023	26	0.00047	28	0.00023
fi10639	520527	20	0.00013	25	0.00008	24	0.00010	28	0.00004	27	0.00005
rl11849	923288	23	0.00019	29	0.00004*	28	0.00006	28	0.00006	25	0.00013
usa13509	19982859	21	0.00011	23	0.00016	24	0.00009	25	0.00007	23	0.00010
xvb13584	(37083)	27	0.00036	29	0.00009	25	0.00045†	25	0.00045†	25	0.00045†
brd14051	469385	25	0.00015	26	0.00009	27	0.00007	29	0.00003*	27	0.00006
mo14185	(427377)	25	0.00005*	23	0.00009	22	0.00013	17	0.00020	24	0.00012
xrb14233	(45462)	9	0.00308	8	0.00330	3	0.00396†	6	0.00374†	12	0.00249
d15112	1573084	15	0.00007	18	0.00005	20	0.00004	18	0.00004	18	0.00003
it16862	557315	6	0.00033	5	0.00027	3	0.00037†	5	0.00032	2	0.00035
xia16928	(52850)	22	0.00082	22	0.00069	23	0.00088	9	0.00227†	25	0.00063
pjh17845	(48092)	11	0.00132	19	0.00083	19	0.00083	19	0.00076*	22	0.00062*
d18512	645238	19	0.00010	21	0.00007	19	0.00014	23	0.00006	24	0.00005
frh19289	(55798)	26	0.00042†	30	0.00000	29	0.00006	27	0.00024†	27	0.00018†
fnc19402	(59287)	22	0.00051	19	0.00062	20	0.00056	14	0.00118	20	0.00056
ido21215	(63517)	19	0.00079	25	0.00026	27	0.00021	23	0.00058	24	0.00031
fma21553	(66527)	23	0.00035*	22	0.00040*	19	0.00065	20	0.00055	26	0.00020*
vm22775	569288	1	0.00107*	2	0.00091*	3	0.00094*	1	0.00119	2	0.00097*
lsb22777	(60977)	23	0.00038	27	0.00016*	26	0.00022	26	0.00022	28	0.00011*
xrh24104	(69294)	28	0.00010	28	0.00010	28	0.00010	29	0.00005	29	0.00005
sw24978	855597	20	0.00020*	14	0.00033	17	0.00020*	12	0.00032	18	0.00023*
Average		19.6	0.00053	21.0	0.00044	20.7	0.00049	19.5	0.00061	21.7	0.00038

Table 2. Solution quality of the GA using the diversity measures H_m and H'_m

	$m=1$		$m=2$		$m=3$		$m=4$		$m=6$		$m=8$	
Div.	#S	A-Err	#S	A-Err	#S	A-Err	#S	A-Err	#S	A-Err	#S	A-Err
H_m	16.7	0.00085	18.1	0.00066	18.2	0.00065	19.1	0.00054	16.1	0.00063	11.8	0.00094
H'_m	16.7	0.00085	19.0	0.00069	19.8	0.00061	20.5	0.00053	21.5	0.00046	20.8	0.00047

Note: When no diversity measure is incorporated, #S and A-Err are 1.2 and 0.00544, respectively.

Next, we focus on the results of H_8^{tr2} with $ratio = 0.1$. We can see that the solution quality of H_8^{tr2} is better than that of H_6^{tr2}. This result also indicates that the core idea of H_m^{tr2} make it possible to successfully capture higher-order dependencies while reducing the problem of the lack of sufficient statistics. Moreover, the use of H_8^{tr2} achieves the best solution quality among all population diversity measures including H'_6.

6 Conclusion

The proposed population diversity measure H_m^{tr1} is defined as the entropy rate of the variable-order Markov process with the aim of capturing higher-order dependencies in sequences of vertices in the population while reducing the problem of the lack of sufficient statistics. The use of this diversity measure, however, has shown only a slight improvement in evaluating population diversity over the previously proposed entropy-based diversity measure H_m, which is based on the fixed-order Markov model. On the other hand, another variant of the proposed population diversity measure H_m^{tr2} has succeeded in improving the abilities of H_m and H_m^{tr1} by further reducing the problem of the lack of sufficient statistics. This research has shown a potential of entropy-based population diversity measures that take into account dependencies between the variables, and the efficacy of the proposed population diversity measures should be investigated on other permutation problems in the future work.

References

1. Begleiter, R., El-Yaniv, R., Yona, G.: On prediction using variable order markov models. J. Artif. Intell. Res. **22**, 385–421 (2004)
2. Colinge, J., Bennett, K.L.: Introduction to computational proteomics. PLoS Comput. Biol. **3**(7), e114 (2007)
3. Maekawa, K., Mori, N., Tamaki, H., Kita, H., Nishikawa, Y.: A genetic solution for the traveling salesman problem by means of a thermodynamical selection rule. In: Proceedings of 3rd IEEE Conference on Evolutionary Computation, pp. 529–534 (1996)
4. Mori, N., Kita, H., Nishikawa, Y.: Adaptation to a changing environment by means of the thermodynamical genetic algorithm. In: Ebeling, W., Rechenberg, I., Voigt, H.-M., Schwefel, H.-P. (eds.) PPSN 1996. LNCS, vol. 1141, pp. 513–522. Springer, Heidelberg (1996)

5. Nagata, Y., Kobayashi, S.: A powerful genetic algorithm using edge assembly crossover for the traveling salesman problem. Informs J. Comput. **25**(2), 346–363 (2013)
6. Nagata, Y., Ono, I.: High-order sequence entropies for measuring population diversity in the traveling salesman problem. In: Middendorf, M., Blum, C. (eds.) EvoCOP 2013. LNCS, vol. 7832, pp. 179–190. Springer, Heidelberg (2013)
7. Tsai, H., Yang, J., Tsai, Y., Kao, C.: An evolutionary algorithm for large traveling salesman problems. IEEE Trans. Syst. Man Cybern. B Cybern. **34**(4), 1718–1729 (2004)
8. Tsujimura, Y., Gen, M.: Entropy-based genetic algorithm for solving tsp. In: Proceedings of 2nd International Conference on Knowledge-Based Intelligent Electronic Systems, pp. 285–290. IEEE (1998)
9. Vallada, E., Ruiz, R.: Genetic algorithms with path relinking for the minimum tardiness permutation flowshop problem. Omega **38**(1), 57–67 (2010)
10. Wang, Y., Lü, Z., Hao, J.-K.: A study of multi-parent crossover operators in a memetic algorithm. In: Schaefer, R., Cotta, C., Kołodziej, J., Rudolph, G. (eds.) PPSN XI. LNCS, vol. 6238, pp. 556–565. Springer, Heidelberg (2010)
11. Zhang, C., Su, S., Chen, J.: Efficient population diversity handling genetic algorithm for QoS-aware web services selection. In: Alexandrov, V.N., Albada, G.D., Sloot, P.M.A., Dongarra, J. (eds.) ICCS 2006. LNCS, vol. 3994, pp. 104–111. Springer, Heidelberg (2006)