Diseño Inteligente de Redes Inalámbricas de Comunicación
Conserjería de Innovación, Ciencia y Empresa
P07-TIC-03044
http://diricom.lcc.uma.es

Málaga, 12 de Diciembre de 2012

Informe Ejecutivo

TÍTULO: THEO-1.0-2012: Cálculo exacto de la Fitness-Distance Correlation usando teoría de landscapes

RESUMEN: La teoría de *landscapes* es un marco formal en el que los problemas de optimización combinatoria se pueden caracterizar como una suma de un tipo especial de *landscape* denominado *landscape elemental*. La descomposición de la función objetivo de un problema en sus componente elementales se puede explotar para calcular estadísticas del problema. En este documento presentamos expresiones en forma cerrada para calcular la *fitness-distances correlation* (FDC) basadas en la descomposición en *landscapes* elementales de problemas definidos para cadenas binarias en las que la función objetivo tiene un solo óptimo global. Presentamos algunos resultados teóricos que cuestionan el uso de FDC como medida de dificultad de un problema.

OBJETIVOS:

1. Presentar una ecuación para calcular FDC en funciones binarias con un solo óptimo global.
2. Proporcionar resultados teóricos que ponen en duda la utilidad de FDC como medida de dificultad de un problema.

CONCLUSIONES:

1. La fórmula exacta proporcionada permite calcular muy eficientemente la FDC sobre problemas binarios. Además permite razonar sobre dicha medida.
2. FDC no parece una buena candidata para medir la dificultad de un problema. El coeficiente de autocorrelación o la longitud de autocorrelación parecen mejores alternativas.

RELACIÓN CON ENTREGABLES: No tiene dependencias.

Diseño Inteligente de Redes Inalámbricas de Comunicación
Conserjería de Innovación, Ciencia y Empresa
P07-TIC-03044
`http://diricom.lcc.uma.es`

Málaga, December 12th, 2012

Executive Summary

TITLE: THEO-1.0-2012: Exact Computation of Fitness-Distance Correlation via Landscape Theory

ABSTRACT: Landscape theory provides a formal framework in which combinatorial optimization problems can be theoretically characterized as a sum of a special kind of landscapes called elementary landscapes. The decomposition of the objective function of a problem into its elementary components can be exploited to compute summary statistics. We present closed-form expressions for the fitness-distance correlation (FDC) based on the elementary landscape decomposition of the problems defined over binary strings in which the objective function has one global optimum. We present some theoretical results that raise some doubts on using FDC as a measure of problem difficulty.

GOALS:

1. Present a closed-form formula to compute FDC on binary problems with one only global optimum.
2. Provide theoretical results that raise some doubts on the usefulness of FDC as a measure of the problem diffculty.

CONCLUSIONS:

1. The closed-form formula provided allows one to compute very efficiently FDC over binary problems. It also allows to take some conclusions on FDC.
2. FDC does not seem a good candidate for a measure of difficulty of a problem. The autocorrelation coefficient or the autocorrelation length seem to be better alternatives.

RELATION WITH DELIVERABLES: No dependencies.

Exact Computation of Fitness-Distance Correlation via Landscape Theory

DIRICOM

December 12^{th}, 2012

1 Reference

Francisco Chicano and Enrique Alba, **Exact Computation of the Fitness-Distance Correlation for Pseudo-boolean Functions with One Global Optimum**, EvoCOP 2012, LNCS 7245, pp. 111-123.

This paper received the **best paper award** of the conference.

2 Summary

The theory of landscapes focuses on the analysis of the structure of the search space that is induced by the combined influences of the objective function of the optimization problem and the neighborhood operator [5]. In the field of combinatorial optimization, this theory has been previously used to characterize optimization problems [3], improve search algorithms [4], and obtain global statistics of the problems [8].

A *landscape* for a combinatorial optimization problem (COP) is a triple (X, N, f), where X is the set of *tentative solutions* of the COP, $f : X \mapsto \mathbb{R}$ defines the objective or *fitness function* and N is the *neighborhood operator*. There exists a special kind of landscapes, called *elementary landscapes* (EL), which are of particular interest due to their properties [9]. Elementary landscapes are characterized by the *Grover's wave equation*:

$$\underset{y \in N(x)}{\operatorname{avg}}\{f(y)\} = \frac{1}{d} \sum_{y \in N(x)} f(y) = f(x) + \frac{\lambda}{d} \left(\bar{f} - f(x)\right), \tag{1}$$

where d is the size of the neighborhood, $|N(x)|$, which we assume is the same for all the solutions in the search space, $\bar{f}$ is the average solution evaluation over the entire search space and λ is a characteristic constant. For a given problem instance whose objective function is elementary, the values $\bar{f}$ and λ can be easily computed in an efficient way, usually from the problem data. Thus, the wave equation makes it possible to compute the average value of the fitness function f evaluated over all of the neighbors of x using only the value $f(x)$, without actually evaluating any of the neighbors.

When the landscape is not elementary it is always possible to write the objective function as a sum of elementary components, called *elementary landscape decomposition* (ELD) of a problem [2]. In the case of binary strings with length n under the one-change neighborhood, the number of elementary components is at most n. Then, Grover's wave equation can be applied to each elementary component and all the results are summed to give the average fitness in the neighborhood of a solution. Furthermore, for some problems the average cannot be limited to the neighborhood of a solution, but it can be extended to the second-order neighrbors (neighbors of neighbors), third-order neighbors, and, in general, to any arbitrary region around a given solution, including the whole search space. Sutton *et al.* [7] show how to compute the averages over *spheres* and *balls* of arbitrary radius around a given solution in polynomial time using the elementary landscape decomposition of pseudoboolean functions.

Landscape theory has been proven to be quite effective computing summary statistics of the optimization problem. Measures like the autocorrelation length and the autocorrelation coefficient can be efficiently computed using the ELD of a problem [3]. Recently, Chicano and Alba [1] and Sutton and Whitley [6] have shown how the expected value of the fitness of a mutated individual can be exactly computed using the ELD. In short, landscape theory can be applied to any COP and thus is generally beneficial for the whole community in discrete optimization, representing a general and usable formalism in practice.

The main contribution of the present work is an exact expression for the Fitness-Distance Correlation (FDC) of COPs defined over a set of binary strings (pseudoboolean functions) having one global optimum. This expression is based on the ELD of the problem. We also analyze the expression in order to dicuss the usefulness of the FDC as a difficulty measure for a problem.

References

[1] Francisco Chicano and Enrique Alba. Exact computation of the expectation curves of the bit-flip mutation using landscapes theory. In *Proceedings of GECCO*, pages 2027–2034, 2011.

[2] Francisco Chicano, L. Darrell Whitley, and Enrique Alba. A methodology to find the elementary landscape decomposition of combinatorial optimization problems. *Evolutionary Computation*, 19(4):597–637, 2011.

[3] Ricardo García-Pelayo and Peter Stadler. Correlation length, isotropy and meta-stable states. *Physica D: Nonlinear Phenomena*, 107(2-4):240–254, Sep 1997.

[4] Guanzhou Lu, Rami Bahsoon, and Xin Yao. Applying elementary landscape analysis to search-based software engineering. In *Proceedings of SSBSE*, pages 3–8, 2010.

[5] P. F. Stadler. Toward a theory of landscapes. In R. López-Peña, R. Capovilla, R. García-Pelayo, H.Waelbroeck, and F. Zertruche, editors, *Complex Systems and Binary Networks*, pages 77–163. Springer, 1995.

[6] Andrew M. Sutton, Darrell Whitley, and Adele E. Howe. Mutation rates of the (1+1)-EA on pseudo-boolean functions of bounded epistasis. In *Proceedings of the 13th annual conference on Genetic and evolutionary computation*, GECCO '11, pages 973–980, 2011.

[7] Andrew M. Sutton, L. Darrell Whitley, and Adele E. Howe. Computing the moments of k-bounded pseudo-boolean functions over Hamming spheres of arbitrary radius in polynomial time. *Theoretical Computer Science*. in press (doi: 10.1016/j.tcs.2011.02.006).

[8] Andrew M. Sutton, L. Darrell Whitley, and Adele E. Howe. A polynomial time computation of the exact correlation structure of k-satisfiability landscapes. In *Proceedings of GECCO*, pages 365–372, New York, NY, USA, 2009. ACM.

[9] Darrell Whitley, Andrew M. Sutton, and Adele E. Howe. Understanding elementary landscapes. In *Proceedings of GECCO*, pages 585–592, New York, NY, USA, 2008. ACM.