Entry for Humies Award 2016 by Kalyanmoy Deb and Christie Myburgh

1. The complete title of one (or more) paper(s) published in the open literature describing the work that the author claims describes a human-competitive result;

Breaking Billion-Variable Barrier in Real-World Resource Allocation Problem Using a Customized Genetic and Evolutionary Algorithm

2. The name, complete physical mailing address, e-mail address, and phone number of EACH author of EACH paper(s);

Kalyanmoy Deb
428 S. Shaw Lane, 2120 EB
Michigan State University
East Lansing, MI 48824, USA
Email: kdeb@egr.msu.edu
Phone: +1 517 930 0846

Christie Myburgh
Principal R&D Engineer
Maptek Pty Ltd
Northbridge, WA 6003, Australia
Email: christie.myburgh@maptek.com.au
Phone: +61 8 6211 0000

3. The name of the corresponding author (i.e., the author to whom notices will be sent concerning the competition);

Kalyanmoy Deb, kdeb@egr.msu.edu

4. The abstract of the paper(s);

Abstract: Despite three decades of intense studies of genetic and evolutionary computation (GEC), researchers outside the EC community still have a general impression that GEC methods are expensive and are not efficient in solving large-scale problems. In this work, we consider a specific integer linear programming (ILP) problem which, although comes from a specific industry, is similar to many other practical resource allocation and assignment problems. Based on a population based evolutionary optimization framework, we develop a computationally fast method to arrive at a near-optimal solution repeatedly. Two popular softwares (*glpk* and *CPLEX*) are not able to handle around 300 and 2,000 integer variable versions of the problem, respectively, even after running for several hours. Our proposed method is able to find a near-optimal solution in less than a second on the same computer. Moreover, the main highlight of this study is that our method scales in a sub-quadratic computational complexity in handling 50,000 to one billion (10^9) variables. We believe that this is the first time such a large-sized real-world constrained problem has ever been handled using any optimization algorithm. The study clearly demonstrates the reasons for such a fast and scale-up application of the proposed method. The work should remain as a successful case study of GEC methods for years to come.

5. A statement stating why the result satisfies the criteria that the contestant claims (see examples of statements of human-competitiveness as a guide to aid in constructing this part of the submission);

Resource allocation problems commonly appear in practice. Our proposed GEC-based optimization algorithm is developed for a casting scheduling problem in which a series of W kg of molten metal needs to be allocated in making numerous castings of different sizes, so that total amount unused metal is minimum. An exactly identical optimization problem occurs for a cutting stock problem, in which different pieces are to be cut from a series of stocks of length L, so that the total amount of waste is minimum. Similar other resource allocation problems exist, such as manpower scheduling, aircraft scheduling, etc. These problems fall under the category of multi-constraint knap-scak problem. Our proposed PILP method is applicable in most such problems without any change in our algorithm. These problems inherently give rise to mixed-integer linear programs, requiring 50,000 or more integer variables for a routine application in an industry.

Existing mixed-integer programming softwares, such as IBM's CPLEX and MATLAB's *intlinprog* or Octave's *glpk*, are industry standard methods for solving such problems, but they cannot handle more than 1,000 variables. Our proposed GA-based PILP method is able to solve 50,000 variable problems repeatedly in less than 10 seconds. The algorithm is applied to a *million*-variable version of the problem and it required only 5 min and solved a staggerring one *billion* variable version of problem in 6.2 days, thereby exhibiting a sub-quadratic performace on a huge range of variable sizes. This is the first time any optimization algorithm has been applied successfully to solve a real-world problem involving billion variables, let alone a billion of integer variables.

(B) The result is equal to or better than a result that was accepted as a new scientific result at the time when it was published in a peer-reviewed scientific journal.

Deb and his student coauthor solved a similar problem in 2003 involving another GA method, which required about 30 minutes to solve a million-variable version of the problem. Our new and improved PILP method requires only 5 min to solve the same problem. No other methods exist to solve the problem having such a large dimension, hence the results of this work stays as the only and competitive result for a million to a billion variable problem ever solved using any optimization method.

(D) The result is publishable in its own right as a new scientific result independent of the fact that the result was mechanically created.

The result is publishable in its own right for various reasons: (i) For the first time, a billion-variable real-world resource allocation problem involving integer variables is solved, (ii) the existing classical optimization algorithms including a few popular commercial softwares are unable to solve more than a 1,000-variable version of the problem, (iii) the algorithm clearly demonstrates the need of a minimum critical population size and a rcombination operator which together is able to scale-up to such a large dimensional problem.

(F) The result is equal to or better than a result that was considered an achievement in its field at the time it was first discovered.

Since the proposed PILP method outperforms classical point-based algorithms including popular commercial softwares by solving many orders of dimension, the result is better than any known result.

(G) The result solves a problem of indisputable difficulty in its field.

The specific resource allocation problem solved in this work is a multi-consraint knap-sack problem and is considered to be a NP-hard problem. This means that to solve the problem to optimality, no polynomial time algorithm is possible. In our application, we have set a target of 99.7% utilization (whereas the maximum possible utlization is 100%, thereby making a maximum of 0.03% deviation from the optimal solution). With this near-optimality target, we have reported a sub-quadratic performance on 50,000 to a Billion variable range, which is remarkable. Thus, in addition to a very fast algorithm, we have also demonstrated that if an approximate solution is acceptable, a polynomial time algorithm to NP-hard problems is possible. This provides a pragmatic solution to such large-scale hard problems.

(H) The result holds its own or wins a regulated competition involving human contestants (in the form of either live human players or human-written computer programs).

The proposed PILP method is compared with existing commercial and public-domain computer softwares (IBM's *CPLEX* and Octave's *glpk*), which are routinely used in solving such mixed-integer programming problems. While these methods can solve only up to 1,000 variable version of the problem and need in industry is to handle 50,000 variables or more, no other competing method exists. Our PILP approach solves 50,000 variable problem in less than 10 seconds and extends millions of variables leading to a billion variable version in a reasonable amount of computational time, making our entry a human-competitive approach.

6. A full citation of the paper (that is, author names; publication date; name of journal, conference, technical report, thesis, book, or book chapter; name of editors, if applicable, of the journal or edited book; publisher name; publisher city; page numbers, if applicable);

Deb, K. and Myburgh, C. (in press). Breaking the Billion-Variable Barrier in Real-World Optimization Using a Customized Evolutionary Algorithm. *Proceedings of Genetic and Evolutionary Computation Conference (GECCO-2016)*, ACM Press. Denver, CO, July 20-24, 2016.

7. A statement either that "any prize money, if any, is to be divided equally among the co-authors" OR a specific percentage breakdown as to how the prize money, if any, is to be divided among the co-authors;

If awarded, authors agree to divide the prize money equally between them.

8. A statement stating why the authors expect that their entry would be the "best,"

We believe the following achievements of this work make it best for the prestigious Humies award:

- For the first time, we solve a Billion variable optimization problem from real-world, describing a resource allocation problem having integer variables and two sets of inequality and equality constraints, which represents various problem domains. Although another billion-variable application is reported in 2006 by Sastry and Goldberg, the problem was an artificial, noisy, one-max problem without any constraints and having Boolean variables only. Moreover, the termination criterion was set to achieve a string with only 51% 1s, far away from the optimal string having 100% 1s.
- This application remains as a triumph for the genetic and evolutionary computation field, as we clearly show that a critical sample size in terms of a population is needed to have a successful application. Moreover, switching the customized recombination operator developed in this work does not find even a feasible solution. These features are key for any GEC algorithm and they distinguish GEC algorithms from point-based classical methods. Applications such as these clearly portray the niche of GEC methods in the whole gamut of optimization problem solving tasks and should attract attention from classical optimization researchers. To critics of GEC methods, this application clearly demonstrates their scalability vis-à-vis other competing methods.
- Due to similarity of the core optimization problem to many other practical large-scale resource allocation problems, the proposed methodology should cater for a variety of different practical problems. The large-scale application demonstrated in this work should allow practitioners solving similar resource allocation problems, to do away with simplifying and employing other dimensionality reduction techniques simply because of lack of a suitable scalable optimization method.

9. An indication of the general type of genetic or evolutionary computation used, such as GA (genetic algorithms), GP (genetic programming), ES (evolution strategies), EP (evolutionary programming), LCS (learning classifier systems), GE (grammatical evolution), GEP (gene expression programming), DE (differential evolution), etc.

A genetic algorithm (GA) is used in this work. GA’s population approach and recombination operation are the main reasons for the success of this work.