The Clever Algorithms Project: Overview*

Jason Brownlee
jasonb@CleverAlgorithms.com
The Clever Algorithms Project
http://www.CleverAlgorithms.com

Abstract

An open problem in the field of Artificial Intelligence is that algorithm descriptions are commonly incomplete, inconsistent, and distributed resulting in variable interpretations of algorithms, undue attrition of algorithms, and ultimately bad science. The Clever Algorithms project is an effort to describe a large number of algorithmic techniques from the the fields of Biologically Inspired Computation and Computational Intelligence in a complete, consistent, and centralized manner such that they are accessible, usable, and understandable. The audience for the algorithm descriptions produced by the project include research scientists, engineers, students, and interested amateurs. The outcome of the project is a compendium of algorithm descriptions that will be released as a book and/or made available online as a website. A methodology for completing the project is proposed involving the incremental completion of discrete standalone algorithm descriptions over an extended period of time, allowing continuous review, revision, and refinement of both the content and the algorithm description template. All files and content for the project are released under a permissive license and can be accessed online from the address http://www.CleverAlgorithms.com. Finally, the project is considered a not-for-profit research project in the vein of open source software projects, and a number of areas for research are suggested.

Keywords: `Clever, Algorithms, Project, Overview, Motivation, Problem, Solution, Audience, Methodlogy, Outcomes`

1   Introduction

The Clever Algorithms Project is concerned with the investigation of algorithms from subfields of Artificial Intelligence and presenting them in a consistent and structured manner. This report provides the touchstone for the projects mission. The project is motivated in Section 2 that reviews the open problem of algorithm communication and the detrimental effect it has on progress in the field. The project is presented in Section 3 as an attempt to address this open problem by providing a set of standardized algorithm descriptions. The objectives of the project and the intended audience are listed, and the projects outcomes and proposed methodology are reviewed. Finally, Section 4 highlights some areas for further consideration in order to ready the effort.

2 Motivation

2.1 Problem

Artificial Intelligence is a large field of study and the description and investigation of algorithmic techniques is a central pursuit in subfields such as Machine Learning, Biologically Inspired Computation, and Computational Intelligence. Communication is a critical open problem in the field of Artificial Intelligence, specifically the description of algorithmic techniques. The fields of Biologically Inspired Computation and Computational Intelligence are of particular concern because these fields are concerned with heuristic procedures based in metaphor, making them susceptible to description my analogy and metaphorical nomenclature.

Open Problem: *The communication of algorithmic techniques in the fields of Biologically Inspired Computation and Computational Intelligence is a difficult open problem.*

- The description of techniques are typically *incomplete*. Many techniques are ambiguously described, partially described, or not described at all.

- The description of techniques are typically *inconsistent*. A given technique may be described using a variety of formal and semi-formal methods that also vary across different techniques limiting the transferability of the skills an audience used to realize a technique (such as mathematics, pseudo code, program code, and narratives). An inconsistent representation for techniques mean that the skills used to understand and internalize one technique may not be transferable to realizing other techniques or even extensions of the same technique.

- The description of techniques are typically *distributed*. The description of data structures, operations, and parameterization of a given technique may span an array of papers, articles, books, and source code published over a number of years, the access of which may be restricted and/or difficult to obtain.

2.2 Effect

For the practitioner, an ill described algorithm may be a frustration where the gaps are filled with intuition and 'best guess'. At the other end of the spectrum, a badly described algorithm may an example of bad science and the failure of the scientific method, where the inability to understand and implement a technique may prevent the replication of results or the investigation and extension of a technique.

General Effect: *Poorly described algorithmic techniques in the fields of Biologically Inspired Computation and Computational Intelligence damage those fields.*

- Poorly described techniques result in *inconsistent interpretation*. A field of study is generally concerned with building a corpus of knowledge, the momentum of which is dependent on a common shared understanding. The so-called diversity-of-understanding provided through the continued reinterpretation of an approach may promote a deeper understanding of a technique, although may cripple forward progress through unnecessary replication of effort. Without a consistent understanding of a technique it cannot be accurately compared, broadly and concurrently investigated, or ultimately make meaningful contributions to the broader field.

- Poorly described techniques result in *undue attrition*. Techniques that can only be realistically implemented by the original author may as well not exist, regardless of related claims of efficiency or efficacy. Practitioners must sift through volumes of papers, articles,

books, and sample source code to formulate a viable interpretation of a given technique. This investment of effort is required for each practitioner and each technique resulting in the vast majority of published approaches disappearing into obscurity. If a practitioner cannot realize a described technique, it will likely be lost as the practitioners move onto those techniques that they can realize with less effort.

- Poorly described techniques result in *bad science*. In a scientific setting, an algorithmic procedure is investigated both theoretically as an abstraction and empirically as an implementation. As such, investigation of an algorithmic technique using the scientific method requires an unambiguous description of the technique. Those algorithms that cannot be clearly communicated, cannot be subjected to broader study, the results cannot be reproduced, and the approach cannot be meaningfully applied, investigated, or extended.

3 Clever Algorithms Project

3.1 Solution

Algorithmic techniques that emerge from the fields of Computational Intelligence and Biologically Inspired Computation are interesting, surprising, and will potentially contribute alternative paradigms of computation to address the limits of more tradition approaches.

Observed Solution: *A strategy to address the open problem of poor technique communication is to present complete algorithm descriptions in a consistent manner in a centralized location.*

Much effort has been put into this solution with the production of many books and libraries of program code. Algorithms from the fields of Computational Intelligence and Biologically Inspired Computation are more nuanced than algorithms from other fields (such as Operations Research for example). They are typically rooted in metaphor and explained via analogy. Although this may contribute to the technical ambiguity of their description, this is also a strength, provoking interest, imagination, and creativity that motivate practitioners.

The audience defines the completeness of an algorithm's description. Different audiences seek out different descriptions with varied levels of rigor, from a theoretician that requires an unambiguous mathematical description, to an engineer who requires an implementable algorithmic procedure, to the algorithm designer interested in abstracting the strategy from the inspiring system.

3.2 Objectives

The Clever Algorithms Project proposes a compendium of algorithm descriptions. The primary objectives of the project are *completeness*, *consistence*, and *centralization*:

- **Completeness**: A well-defined template shall be defined for describing an algorithm to a selected audience, each section of which will have a clear intention. Algorithms conforming to the template will be considered complete, and all algorithm descriptions listed in the compendium will conform to the proposed template.

- **Consistency**: The conformation of a collection of algorithm descriptions to a template will consider the descriptions consistent, and all algorithm descriptions listed in the compendium shall conform to the same template.

- **Centralization**: Algorithm descriptions shall be accessible via multiple means, although will be managed at a central point of dissemination. The audience of the compendium will consume its content from a centralized access point.

The secondary objectives of the project are *accessibility*, *usability*, and *understandability*:

- **Accessibility**: The algorithm descriptions shall be widely accessible by the intended target audience both physically (such as electronically online and tangibly printed) and practically (such as loosely coupled descriptions amenable to electronic search and hoc access).

- **Usability**: The algorithm descriptions shall be directly usable by the intended audience. The use of an algorithm description will be defined by a specific use case of a specific target audience. Each section in the proposed template will address at least one specific use case of a specific target audience.

- **Understandability**: The algorithm descriptions shall be in a structure and nomenclature suitable to be understood by the target audience and where appropriate written in the English language. Specifically, American English which is the dialect and language of science and technology.

3.3 Audience

The audience for the The Clever Algorithms Project are practitioners concerned or interested with the fields of Computational Intelligence and Biologically Inspired Computation, such as:

- **Scientists**: Research scientists concerned with theoretically or empirically investigating algorithms, addressing questions such as: *What is the motivating system and strategy for a given technique? What are some algorithms that may be used in a comparison within a given subfield or across subfields?*

- **Engineers**: Programmers and developers concerned with implementing, applying, or maintaining algorithms, addressing questions such as: *What is the algorithm procedure for a given technique? What are the best practice heuristics for employing a given technique?*

- **Students**: Undergraduate and graduate students interested in learning about techniques, addressing questions such as: *What are some interesting algorithms to study? How to implement a given approach?*

- **Amateurs**: Practitioners interested in knowing more about algorithms, addressing questions such as: *What classes of techniques exist and what algorithms do they provide? How to conceptualize the computation of a technique?*

3.4 Outcomes

An important objective of the project is the centralization of the prepared algorithm descriptions as a compendium. This will be realized as a deliverable in either, or both a book and website.

- **Book**: Publish the compendium of algorithm descriptions in a book format as an eBook and/or a dead tree book. The preparation of a book would require suitable front and back matter and one or more professional editors if the book is to be published commercially. Self-service publishing (publish on demand) services may be adopted to reduce costs.

- **Website**: Publish the compendium of algorithm descriptions as a website with a tailor made content management system. The preparation of the website would require hosting and careful attention to online marketing, algorithm discoverability, and potentially monetization to cover ongoing hosting costs. Blogs and free hosting solutions may be adopted to reduce costs.

3.5 Methodology

The project will be executed through the completion of many small (2-4 man days), discrete, semi-independent technical reports. Each report will cover a specific aspect of the project, such as the algorithm description template, the process for selecting algorithms to include, and a report for each given algorithm. This methodology for content development facilitates the continued ad hoc contribution to the project through the creation of formal work product. Each report represents a task toward the completion of the project, the discreteness and semi-independence of which may allow a distribution of effort over a small team, if available.

The content of the technical reports may be reviewed, edited, and refined relatively independently. The content of the completed technical reports may then be collected and included and/or adapted into a form suitable for consumption by the target audience, such as book and website formats. The development of the structure and content of the project outcomes will be developed in concert with the production of technical reports.

3.6 Location

The project structure and content is managed via a version control system hosted online currently located at the URL http://www.CleverAlgorithms.com. The managed project supports both content development, editing and revisions, as well as sample program source code.

4 Conclusions

The Clever Algorithms project is an ambitious undertaking that may or may not have a market in the target audience. It is genuinely believed that there is a need and want for a modestly sized corpus of consistently described algorithmic techniques from the fields of Computational Intelligence and Biologically Inspired Computation. The project is a research effort that does not require or expect explicit monetary payoff in the spirit of open source software projects.

Before the long process of preparing descriptions of algorithmic techniques begins, there are some important foundational areas that must be investigated in order to suitably prepare. Some examples of these areas include: the definition of the algorithm description template, a summary of the books and works that inspired the project, a definition of the algorithm selection criteria and process, and a general introduction into the broader fields of study.