

Clever Algorithms: Machine Learning Project Overview*

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

February 7, 2011
Technical Report: CA-TR-20110207-1

Abstract

Machine Learning is a huge field comprised of many tens to hundreds of algorithms, the application of which drive a large number of modern businesses and organizations. The recently published ‘Clever Algorithms: Nature-Inspired Programming Recipes’ has motivated a search for other fields where the standardized algorithm description template might provide a useful tool to improve algorithm communication. This report suggests a follow-up project titled ‘Clever Algorithms: Machine Learning Programming Recipes’ that applies the previously adopted algorithm template to the algorithms and methods from the field of Machine Learning, the results of which are also intended to be published as a free website and print-on-demand book.

Keywords: Clever Algorithms, Machine Learning, Project Overview

1 Introduction

The previous book project “Clever Algorithms: Nature-Inspired Programming Recipes” has been completed and the book launched successfully [3]. As such thoughts and efforts have turned toward a new project. This document outlines the proposal for a follow-up book that follows the same structure and methodology as the ‘Nature-Inspired’ book, although focuses on the related field of Machine Learning algorithms.

There is a lot of overlap between the algorithms in the previous book and so-called machine learning methods. In fact, some algorithms from the fields of Metaheuristics, Computational Intelligence, and Biologically Inspired Computation may be considered Machine Learning algorithms. The features used to differentiate this project from the previous are as follows:

- *Basis*: Algorithms in the ‘Nature-Inspired’ book were selected because of their biological or natural analogy or metaphor, or because they were optimization methods associated with such algorithms from the general fields of Metaheuristics, Computational Intelligence, and Biologically Inspired Computation (such as the stochastic algorithms described in the book). Machine Learning methods generally do not have a biological or natural inspiration and have a statistical basis.
- *Models*: Machine Learning algorithms are generally concerned with the development or construction of a model based on empirical data (provided or collected), where as the

*© Copyright 2011 Jason Brownlee. Some Rights Reserved. This work is licensed under a Creative Commons Attribution-Noncommercial-Share Alike 2.5 Australia License.

majority of the algorithms described in the ‘Nature-Inspired’ text were mostly optimization methods that sought a combination of parameters that resulted in a minimum or maximum of a cost function.

This report reviews the general open problem in Artificial Intelligence of algorithm communication (2), and goes on to describe the objectives, audience and outcomes of the proposed Clever Algorithms: Machine Learning Programming Recipes project (3).

2 Problem

The open problem of technique description was articulated well in the project overview for the previous work [1]. The central argument contended that algorithm descriptions were generally incomplete, inconsistent, and distributed, and this resulted in the inconsistent interpretation of algorithms, undue attrition of methods, and the increased chance of bad science.

This assessment applies to so-named nature-inspired algorithms of the first book drawn from the fields of Metaheuristics, Computational Intelligence, and Biologically Inspired Computation. This assessment is believed to also apply to the field of Machine Learning algorithms, although to a lesser degree. It is believed that Machine Learning algorithms are generally described more succinctly using mathematical methods, although are not completely precluded from suffering from this open problem of communication.

3 Clever Algorithms: Machine Learning Programming Recipes

3.1 Solution

The algorithmic methods described in the field of Machine Learning are diverse, interesting, and drive many modern businesses and decision making processes.

Many books have been published on the field of Machine Learning and its methods. Additionally, many excellent academic and professional level libraries have been created to present standardized and executable versions of popular Machine Learning algorithms, many of them free and open source.

That being said, a concise reference resource that uses a standardized algorithm template such as that developed for the first Clever Algorithms ([2]) book does not exist. This highly-structured template, that includes a minimal although executable version of the algorithm itself, is the primary contribution of the first book, and is believed to provide a valuable tool in being reappraised for the presentation of Machine Learning algorithms.

Some Machine Learning algorithm descriptions can be too verbose, leaning on extended descriptions of mechanisms and features, whereas others can be too terse, showing only the central equations to a method. The application of the demonstrated algorithm description template is expected to aid in better communicating a dearth of common and popular Machine Learning methods.

3.2 Objectives

As with the previous project, this project proposes a compendium of algorithm descriptions the primary objectives of which 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 ad 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 project are practitioners concerned or interested with the fields of Artificial Intelligence and Machine Learning, 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 terms of a book and website.

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

- **Website:** A published 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.

The already published ‘Nature-Inspired’ project as a print-on-demand book and website provides a model for the outcomes of this project.

3.5 Methodology

The project will be executed through the completion of many small (2-to-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.

4 Conclusions

Like the ‘Nature-Inspired’ project before it, this is an ambitious project that will require a significant amount of hard work to complete. And like the previous project, it is strongly believed that there is a clear need for the resulting algorithm descriptions to exist. The project is a research effort that does not require or expect explicit monetary payoff in the spirit of open source software projects, although support by subject matter experts, technical and copy editors, and general advocates and evangelists would be warmly received.

5 Contribute

Found a typo in the content or a bug in the source code? Are you an expert in this field and know some facts that could improve the algorithm description for all? Do you want to get that warm feeling from contributing to an open source project? Do you want to see your name as an acknowledgment in print?

Two pillars of this effort are 1) that the best domain experts are people outside of the project, and 2) that this work is subjected to continuous improvement. Please help to make this work less wrong by emailing the author ‘Jason Brownlee’ at jasonb@CleverAlgorithms.com or visit the project website at <http://www.CleverAlgorithms.com>.

References

- [1] J. Brownlee. The clever algorithms project: Overview. Technical Report CA-TR-20100105-1, The Clever Algorithms Project <http://www.CleverAlgorithms.com>, January 2010.
- [2] J. Brownlee. A template for standardized algorithm descriptions. Technical Report CA-TR-20100107-1, The Clever Algorithms Project <http://www.CleverAlgorithms.com>, January 2010.

- [3] Jason Brownlee. *Clever Algorithms: Nature-Inspired Programming Recipes*. Lulu, 2011.