

Extended Abstract

Feature Selection with Limited Bit Depth Mutual Information for Embedded Systems [†]

Laura Morán-Fernández *, Verónica Bolón-Canedo and Amparo Alonso-Betanzos

CITIC, Universidade da Coruña, 15071 A Coruña, Spain; vbolon@udc.es (V.B.-C.); ciamparo@udc.es (A.A.-B.)

* Correspondence: laura.moranf@udc.es

† Presented at the XoveTIC Congress, A Coruña, Spain, 27–28 September 2018.

Published: 17 September 2018

Abstract: Data is growing at an unprecedented pace. With the variety, speed and volume of data flowing through networks and databases, newer approaches based on machine learning are required. But what is really big in Big Data? Should it depend on the numerical representation of the machine? Since portable embedded systems have been growing in importance, there is also increased interest in implementing machine learning algorithms with a limited number of bits. Not only learning, also feature selection, most of the times a mandatory preprocessing step in machine learning, is often constrained by the available computational resources. In this work, we consider mutual information—one of the most common measures of dependence used in feature selection algorithms—with reduced precision parameters.

Keywords: feature selection; mutual information; reduced precision; embedded systems; Big Data

1. Introduction

In the age of Big Data, with datasets being collected in almost all fields of human endeavor, there is an emerging economic and scientific need to extract useful information from it. Thus, machine learning algorithms have become indispensable. One machine learning technique is feature selection [1]. It arises from the need of determining the “best” subset of variables for a given problem. The use of an adequate feature selection method can avoid over-fitting and improve model performance, providing faster and more cost-effective learning models and a deeper insight into the underlying processes that generate the data. Features can be categorized in three ways: relevant, irrelevant and redundant. As a result, selecting the relevant features and ignoring the irrelevant and redundant ones is advisable.

The process of feature selection is typically performed on a machine using high numerical representation (64 bits). Using a more powerful processor provides significant benefits in terms of speed and capability to solve more complex problems. Although this capability does not come without cost; a conventional microprocessor can require a substantial amount of off-chip support hardware, memory, and often a complex operating system. In contrast to up-to-date computers, these requirements are often not met by embedded systems, low energy computers or integrated solutions that need to optimize the used hardware resources. With the power demand of smartphones, health wearables and fitness trackers, there is a need for tools that enable energy consumption estimation for such systems. Thus, we identify one such opportunity to develop a feature selection algorithm in embedded systems without reducing performance. This opportunity leverages the observation that algorithms yield parameters which can achieve performances close to that of optimal double-precision parameters by simply limiting the amount of bits. In this work, we investigate feature selection by considering the information theoretic measure of mutual information with reduced precision parameters. The mutual information measure is used due to its computational efficiency

and simple interpretation. Therefore, we are able to provide a limited bit depth mutual information, and, through minimum Redundancy Maximum Relevance feature selection method, experimentally achieve classification performances close to that of 64-bit representations for several real and synthetic datasets.

2. Limited Bit Depth Mutual Information

In information theoretic feature selection, the main challenge is to estimate the mutual information [2]. To calculate mutual information we need to estimate the probability distributions. Internally, it counts the occurrences of values within a particular group. Thus, based on Tschitschek's work [3] for approximately computing probabilities, we investigate mutual information with limited number of bits by considering this measure with reduced precision counters. To perform the reduced precision approach, we target a fixed-point representation instead of the 64-bit resolution.

Mutual Information parameters are typically represented in the logarithm domain. For the reduced precision parameters, we compute the number of occurrences and use a lookup table to determine the logarithm of the probability of a particular event. The lookup table is indexed in terms of number of occurrences of an event and the total number of events and stores values for the logarithms in the desired reduced precision representation. Following the fixed-point representation, and to limit the maximum size of the lookup table and the bit-width required for the counters, we assumed some maximum integer number. After calculating the cumulative count, in order to guarantee that the counts stay in range, the algorithm identifies counters that reach their maximum value, and halves these counters.

3. Experimental Results and Conclusions

Our limited depth mutual information can be applied to any method that uses internally the mutual information measure. We have chosen to do it within feature selection since with the advent of Big Data, feature selection process has a key role to play in helping reduce high-dimensionality in machine learning problems. There is a large number of feature selection methods that use mutual information as a measure, thus their performance depending on the accuracy obtained by the mutual information step. Among the different feature selection algorithms based on mutual information, the mRMR (minimum Redundancy Maximum Relevance) multivariate filter [4] is used due to its popularity and good results in the machine learning area.

Experimental results over several synthetic and real datasets have shown that 16 bits are sufficient to return the same feature ranking than that of double precision representation. Besides, classification results showed that even using a 4-bit representation, our limited bit depth mutual information was able to achieve performances very close to that of full precision mutual information. As a result, meaningful computational, runtime and memory benefits will be provided when implementing mutual information in embedded systems.

Acknowledgments: This research has been financially supported in part by the Spanish Ministerio de Economía y Competitividad (research project TIN2015-65069-C2-1-R), by European Union FEDER funds and by the Consellería de Industria of the Xunta de Galicia (research project GRC2014/035). Financial support from the Xunta de Galicia (Centro singular de investigación de Galicia accreditation 2016–2019) and the European Union (European Regional Development Fund–ERDF), is gratefully acknowledged (research project ED431G/01).

Conflicts of Interest: The author declares no conflict of interest.

References

1. Guyon, I.; Elisseeff, A. An introduction to variable and feature selection. *J. Mach. Learn. Res.* **2003**, *3*, 1157–1182.
2. Shannon, C.E. A mathematical theory of communication. *ACM SIGMOBILE Mob. Comp. Rev.* **2001**, *5*, 3–55.

3. Tschatschek, S.; Pernkopf, F. Parameter learning of Bayesian network classifiers under computational constraints. In Proceedings of the Joint European Conference on Machine Learning and Knowledge Discovery in Databases, Porto, Portugal, 7–11 September 2015; Springer, pp. 86–101.
4. Peng, H.; Long, F.; Ding, C. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Trans. Pattern Anal.* **2005**, *27*, 1226–1238.



© 2018 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).