ESANN 2014 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 23-25 April 2014, i6doc.com publ., ISBN 978-287419095-7.
Available from http://www.i6doc.com/fr/livre/?GCOI=28001100432440.

Context- and cost-aware feature selection in ultra-low-power sensor interfaces

Steven Lauwereins[1], Komail Badami[1], Wannes Meert[2], Marian Verhelst[1]

1- KU Leuven - ESAT-MICAS
Kasteelpark Arenberg 10, Heverlee B-3001 - Belgium

2- KU Leuven - CS-DTAI
Celestijnenlaan 200A, Heverlee B-3001 - Belgium

Abstract. This paper introduces the use of machine learning to improve efficiency of ultra-low-power sensor interfaces. Adaptive feature extraction circuits are assisted by hardware embedded learning to dynamically activate only most relevant features. This selection is done in a context and power cost-aware way, through modification of the C4.5 algorithm. Furthermore, context dependence of different feature sets is explained. As proof-of-principle, a Voice Activity Detector is expanded with the proposed context- and cost-dependent voice/noise classifier, resulting in an average circuit power savings of 75%, with negligible accuracy loss.

1 Introduction

Due to the increasing number of sensors integrated on a variety of ubiquitous electronic devices, the boundaries between the domains of machine learning, circuit design and signal processing are blurring more than ever [1]. Namely, extracting information from these always-on sensors is strongly constrained by the limited power availability in the sensing devices. To be able to deal with the sensory data overload in a power-efficient way, it is important to discard irrelevant data as soon as possible, and extract only information-carrying features. It has, for example, been well established that processing the raw data on-board is far more cost-effective[1] compared to transmitting raw data to a central data collecting node [2]. Taking this one step further, also within a sensor node relevant features should be extracted as close to the raw sensor as possible to avoid power wastage. Hence, adaptive ultra-low power consumption chips, such as our work in [3], demand an intelligent selection mechanism to pick essential features for information extraction, allowing the system to selectively enable and disable particular hardware feature extraction blocks. This paper will propose machine learning algorithms to enable such optimal hardware activation in future sensor interfaces.

In this regard, Section 2 will introduce the notion of context-aware feature activation, to dynamically activate current most relevant features. Section 3 further expands this idea to context-aware and feature cost-aware classification for improved resource efficiency, taking also feature circuit power cost into account. Section 4 applies the derived approach on a proof-of-principle hardware design of

[1] unless explicitly mentioned, in this paper cost refers to power consumption

ESANN 2014 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 23-25 April 2014, i6doc.com publ., ISBN 978-287419095-7.
Available from http://www.i6doc.com/fr/livre/?GCOI=28001100432440.

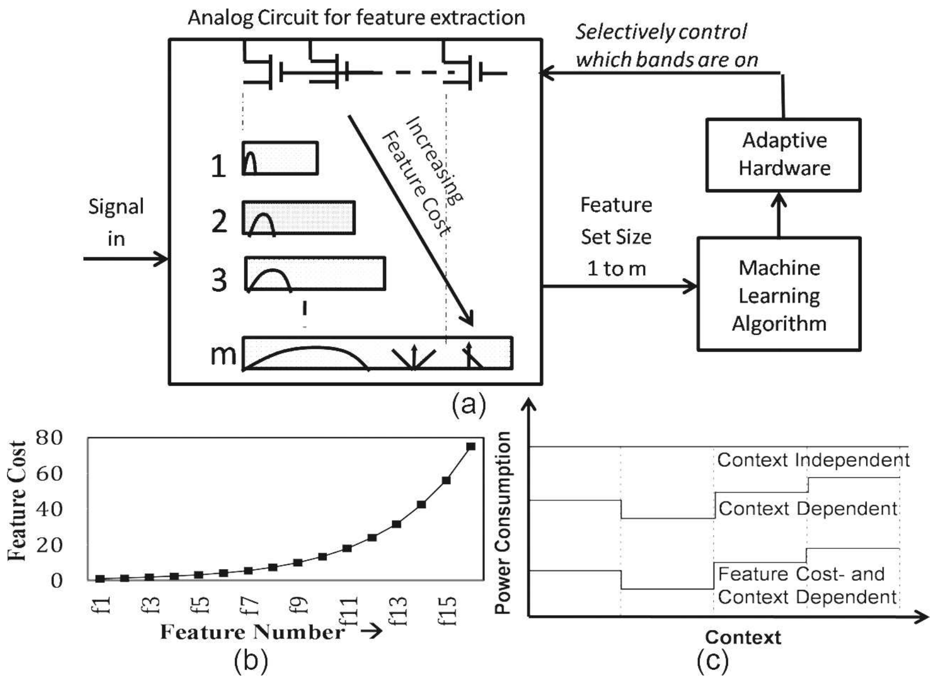

Fig. 1: (a) Context- and Feature cost-aware classifier showing integration of adaptive hardware and machine learning. (b) Illustrates the varying power cost across feature extraction hardware blocks. (c) Show anticipated power consumption versus context for three types of classifiers. (All applicable for acoustic sensing scenario introduced in section 4)

a Voice Activity Detector, demonstrating 4X reduction in power consumption. It also uncovers the context dependency of different feature sets on tree cost.

2 Context-aware feature selection

In many applications, the relative information content of a feature is highly context-dependent. Depending on the context, some features port a more distinctive value towards the classes of interest. Examples are acoustic classifiers, prone to various types of background noises, or patient-specific biomedical data classification. To always operate the target sensing systems at maximal resource efficiency, feature selection should be done in a context-aware way (See Fig. 1a). In this work, decision tree based classifiers will be trained for different contexts with only discriminative features being selected per context. As will be demonstrated in Section 4, this context-awareness allows cutting the number of actively observed features by a factor of 1.6X in typical sensing applications. To preserve power savings, the overhead of the introduced intelligence should be low. A tree based classifier is chosen for its amenability to low-footprint hardware implementation, comprising a series of comparisons. The decision tree automatically

ESANN 2014 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 23-25 April 2014, i6doc.com publ., ISBN 978-287419095-7. Available from http://www.i6doc.com/fr/livre/?GCOI=28001100432440.

orders the features in order of decreasing information gain, allowing dynamically trading-off the classification accuracy with the power consumption.

- If all labels are identical, create leaf node.
- For each feature 'f_i'
  - Find the normalized information-gain/watt from splitting on 'f_i'
- Let 'f_{best}' be the feature with the highest normalized information gain/watt
- If sufficient gain, create a decision node that splits on 'f_{best}', else create a leaf node
- Recurse on the sublists obtained by splitting on 'f_{best}', and add those nodes as children of node

Algorithm 1: Power cost-aware modification to C4.5

3 Cost- and Context-aware feature selection

In many applications, the computational cost (c_i) to obtain a feature (f_i) is not constant across available features as depicted in Fig 1b. The decision tree would learn the model only based on the information gain of the feature f_i, completely oblivious to the feature cost c_i. While state-of-the-art research, such as [4], introduced cost-aware learning for decision trees, cost optimization is primarily focused mainly on misclassification cost, while the power impact of feature deactivation is not taken into account to the author's knowledge. Also, the algorithm proposed in [4] does not yet support continuous attributes. Another popular technique [5] to save cost is to optimize cost while traversing of the tree. In this case however, this has no impact because the analog circuits have to be (de-)activated beforehand to capture sound. Also, the proposed algorithm is currently only for discrete attributes and, additionally, we cannot optimize the tests while traversing the tree since the analog circuit has to be (de-)activated beforehand. Hence in very lower power applications it is of paramount importance to sensitize the machine learning algorithm to feature cost as well, leading to enabling power consumption profile improvements as shown in Fig. 1c. We suggest the following modification to C4.5 as in pseudo code in Alg. 1. As such, instead selecting a feature with maximum Information-gain, the Information-gain/Watt is used as a metric independent of the previous usage of the feature. This allows the algorithm to relatively lower the rank of features with high feature cost. In other words, this modification gives "best value for money" or highest resource efficiency.

4 Proof-of-principle - Voice Activity Detector

This section will apply the proposed approach to the design of a relevant power-scarce sensor interface to illustrate its merits in a realistic design. An acoustic interface for Voice Activity Detection (VAD) in mobile devices, targeting continuous voice vs. noise classification, is extremely power-constrained due to

Feature Set 1	Feature Set 2	Feature Set 3
Energy	Δ_{Energy}	MSB normalized energy
E_i	$E_i - E_{(i-1)}$	$E >> log_2(E_{total})$

Table 1: Degradation of classification accuracy caused by SNR variation is avoided by using relative features (i.e. Δ & MSB) instead of absolute features.

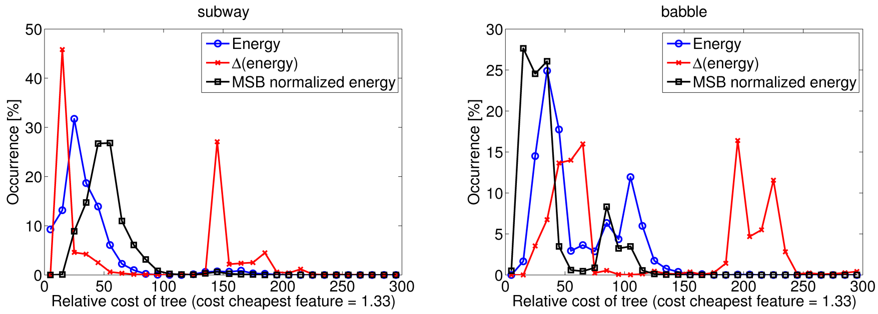

Fig. 2: Context dependence of tree cost (sum cost of all used features) for different feature sets (occurrence in 5000 trees). White noise like contexts are inexpensively classified by Feature set 2 while speech like noise is best classified by Feature set 3. Training SNR from 0 to 20dB; accuracy = 82%

its always-on operation. This design will serve as a proof-of-principle for both context-aware (section 4.1), context-, cost- and feature set-aware (section 4.2) voice/noise classification. Power savings of 1.6X and 4X respectively over context independent operation will be demonstrated.
Fig. 1a shows the implemented VAD system design, where an incoming voice/noise signal is processed in the analog domain by decomposing it into 16 frequency bands spaced on a logarithmic scale. Feature f_i is defined as the mean energy level (E_i) in each of these bands over a time period 100ms. The actual feature cost is determined by executing analog circuit simulations of the filter bank, resulting in strongly rising feature power cost for higher increasing frequency bands numbers, as shown in Fig. 1b.
All results reported in this paper have been conducted using the widely used NOIZEUS dataset [6] for voice and noise samples. Noise and voice samples are combined for varying signal-to-background-noise ratio (SNR) values and context, using MATLAB. A balanced dataset of 2183 samples each for voice and noise is used. Decision tree learning and evaluation is done deploying the modified C4.5 algorithm (J48 using Weka).

4.1 Context-aware Voice/Noise classifier

Within different NOIZEUS contexts, a voice/noise classifier is learned in a supervised way. To ensure real life applicability, one common model is trained across a wide SNR-range, spanning from 0dB to +20dB. Feature E_i is highly SNR depen-

ESANN 2014 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 23-25 April 2014, i6doc.com publ., ISBN 978-287419095-7.
Available from http://www.i6doc.com/fr/livre/?GCOI=28001100432440.

Feature selection for context independent scenario

Context	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	Power
NA	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	382

Feature selection for context dependent scenario

Context	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	Power
Street		■					■		■	■						■	218
Exhibition							■		■		■	■			■	■	308
Subway							■								■	■	236
Babble		■		■			■	■								■	201
Average																	240.75

Feature selection for context and cost dependent scenario

Context	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	Power
Street		■					■			■		■				■	162
Exhibition							■	■	■	■	■						133
Subway					■		■										35
Babble		■			■		■	■	■	■							65
Average																	98.75

Fig. 3: Selected features and power implications for Context / Cost (In)dependent scenarios

dent and hence to avoid accuracy loss due to SNR change, two context dependent and SNR normalized feature sets are proposed in Table 1. Feature set 2 uses difference between the signal levels in adjacent frequency bands whereas Feature set 3 involving SNR normalization by division operation. However since the exact computation of division is power hungry in embedded systems, $log_2(E_{total})$ operation is used to approximating the division using shifting operation. As seen in Fig. 2 the optimal feature set (minimal tree cost) is highly context dependent. For example, contexts with high degree of stationary noise would perform well with Feature set 2 where the differencing operation enhances the spectral features for voice classification and the Feature set 3 appears to be more suited to contexts where noise spectrum highly resembles that of voice. Going forward, the results are evaluated for Feature set 2. The learned context-dependent model shows that not all features are used in the classification and power is only spent on discriminative features. This can be seen in Fig. 3 where depending on the noise-context, only a subset of the features are used in voice/noise classification leading to 1.6X savings in power consumption as compared to a context independent scenario where all the bands are used. Of course, context detection adds a cost penalty but contexts usually do-not change very often and is therefore not as expensive as voice recognition which runs continuously [7]. In this work we don't go into detail how the context is detected but for a cell phone this e.g. can be achieved by combining observed cellular IDs and inertial sensors [8].

4.2 Feature cost- and context-aware Voice/Noise classifier

Cost-awareness was introduced into J48 in Weka, to base classification on Information gain/Watt rather than Information-gain alone as highlighted in Section 3. Rerunning model training, the impact of this modification can be seen in Fig. 3 where, as compared to context dependent scenario for the same context, relatively lower frequency bands are selected for voice/noise classification leading to 4X power savings compared to the context independent case. The analog circuit for feature extraction is designed such that the computation of un-necessary fea-

ESANN 2014 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 23-25 April 2014, i6doc.com publ., ISBN 978-287419095-7. Available from http://www.i6doc.com/fr/livre/?GCOI=28001100432440.

Training (0 to 20dB)	Context Inde-pendent	Context Dependent					Feature cost and context dependent				
		Street	Exhibi-tion	Sub-way	Babble	Mean	Street	Exhibi-tion	Sub-way	Babble	Mean
Testing @ -10dB	64	69	68	72	68	69.25	69	68	68	67	68
Testing @ 0dB	85	75	76	79	74	76	75	75	74	73	74.25
Testing @ 10dB	88	83	85	87	81	84	83	83	81	81	82
Tree size	63	6	8	7	7	7	7	5	3	5	5
Power Cost	382	218	308	236	201	240.75	162	133	35	65	98.75

Fig. 4: Classification accuracy for Context / Cost (In)dependent scenarios.

tures can be deactivated. As shown in Fig. 4, there is no significant difference in the average accuracy numbers between context aware and feature cost and context aware to context independent classification.

5 Conclusions & future work

This paper enables hardware-embedded machine learning to smartly control power-aware adaptivity in future sensor interfaces. Through the introduction of context-aware feature sets and cost-aware decision trees, hardware resource efficiency is maximized by controlled feature (de)activation. A VAD proof-of-principle demonstrates up to 4X power savings in real sensor interfaces, a gain invaluable towards ubiquitous sensing in the internet-of-things. We are currently working towards the chip tape-out of this self-adaptive context-aware voice activity detector. Furthermore, the C4.5 algorithm will be expanded with a dynamic global tree cost function, capable of keeping track of which frequency bands were already activated, and capable of a context-aware selection of the best feature set in Table 1 without performance loss.

References

[1] Hsu, Shu-Yu, et al., "A 48.6-to-105.2uW machine-learning assisted cardiac sensor SoC for mobile healthcare monitoring.", *VLSI Circuits (VLSIC), 2013 Symposium on. IEEE, 2013* (VLSIC 2013).

[2] Calhoun, Benton H., et al., "Body sensor networks: A holistic approach from silicon to users." *Proceedings of the IEEE 100.1 (2012)*, pages 91-106, 2012.

[3] Verhelst, Marian, et al., "A reconfigurable 0.13um CMOS 110pJ/pulse, fully integrated IR-UWB receiver for communication and sub-cm ranging." *Solid-State Circuits Conference-Digest of Technical Papers, 2009. ISSCC 2009. IEEE International.*, IEEE, 2009.

[4] Ling, Charles X., et al., "Decision trees with minimal costs." *Proceedings of the twenty-first international conference on Machine learning.* ACM, 2004.

[5] Xu, Z., et al., "Cost-Sensitive Tree of Classifiers." *International Conference on Machine Learning*, 2013.

[6] Hu, Y., Loizou, P., "NOIZEUS: A noisy speech corpus for evaluation of speech enhancement algorithms." 2005.

[7] Siewioreck, D., et al., "SenSay: A context-Aware Mobile Phone.", page 248, *Seventh IEEE International Symposium on Wearable Computers (ISWC'03)*,2003.

[8] Jiang, Yifei, et al., "User-Centric Indoor Air Quality Monitoring on Mobile Devices." *AI Magazine 34.2:11*, 2013.