Active Machine Learning of Complex Visual Tasks

Phil Sheridan
School of Computing and Information Technology
Griffith University, Brisbane, Australia
p.sheridan@griffith.edu.au

Steve Drew
School of Computing and Information Technology
Griffith University, Brisbane, Australia
s.drew@griffith.edu.au

Abstract

This paper reports on the development of an artificial vision system implemented in software and its application to mammography. It describes a supervision strategy that facilitates the machine-centered learning of complex visual tasks. The key contributions of this paper are the description of our "active" learning strategy and a mechanism by which pixels associated with individual artifacts visible to a human eye in an image can be captured and used as training examples for a machine-learning algorithm. Techniques are discussed in the context of the analysis of micro-calcifications. The significance is that it provides a means by which ill-defined concepts (e.g. visual characteristics of tumors) that are embedded in a complex image (e.g. mammograms) can be more efficiently and accurately learned by a machine.

Keywords

Machine vision, hexagonal lattice, automated mammography, space-variant sensor

1. INTRODUCTION

Breast cancer is the most common form of cancer in women and the second highest cause of death for women in the world . One million new cases were discovered last year with over 580,000 of those coming from the United States, Europe and Australia. Between one third and one half of that number of cases currently add to the mortality total each year [1,2]. Consequently these same countries are leading the research into breast cancer detection and treatment.

A leveling of the rate of mortality and morbidity due to breast cancer in western countries has been attributed to the various programs of early detection and intervention [3]. This enables most cancers can be detected while still relatively small and more successfully treatable. With some qualification, [4-10] screening mammography is considered the best early detection method available. Consequently, most national guidelines recommend a combination of procedures including periodic clinical examination and screening mammography for women over the age of 40 years [3].

Screening mammography is typified by a huge volume of cases (sets of radiographs) to be processed with a very low yield of detectable abnormalities. Correctly and consistently detecting and diagnosing early stages of masses and micro-calcification clusters from the range of complex "normal" background breast tissue arrangements has proven to be a difficult, tedious and time-hungry task for most mammography radiologists [4,5].

With low intrinsic specificity, one feature of current CAD applications is that as the sensitivity is increased the number of false positive indications also increases, leading to increased patient recall rate. Conversely as sensitivity is decreased then the number of false negative indications increases, meaning that more tumor indications are missed [4]. At this time, no CAD system can approach the optimal combination of sensitivity and specificity that a competent screening radiologist can attain [11]. Sensitivity in most CAD tests is acceptable but the best figures for specificity are less than one third of a radiologist practiced in screening mammography.

It appears that before any confident reduction of their workload with CAD can happen the specificity figures must improve dramatically. In essence this is a problem of expanding the capabilities of machine vision and learning with respect to digital image analysis.

From a graphical analysis perspective, discerning indications of cancer from the complex background of breast parenchyma is essentially a "signal to noise" exercise [4]. A trained radiologist can classify more than a dozen different abnormal tissue artifacts from an infinite range of normal tissue densities and arrangements. Each type of artifact might appear in countless different configurations, ensuring that program-driven machine learning, concept generalization and classification remains unachieved.

This paper reports on the in-progress development of a software-based machine vision/learning system named "Akamai". The word "Akamai" comes from the Hawaiian language and means "smart" or "intelligent". Akamai presents a human-supervised machine learning process that captures expert knowledge using image mark-up tools, to train the machine to visually recognize and classify image artifacts in digital mammograms. Using this software system, the machine learner is trained to "see" what the expert sees and correlate this with the expert's determination of the detected image artifact.

Sufficient, selected training examples with significant features indicated, allows us to create a learner that can generalize a concept from accumulated knowledge and apply it to the task of classification. In Akamai, a "lazy" or supervisor-centered learning mode with the highest level of human supervision, each training example might take the expert several minutes to load, mark-up and classify. With a complex concept, requiring a large number of training examples, the supervision overhead soon becomes prohibitive.

We describe here a progressive machine learning approach that is learner-centered and allows the machine to take advantage of its increasing "expertise" to minimize human supervisor input. A sequence of increasingly machine-centered learning modes move the machine from a slow, "passive" learner to one that is actively and interactively seeking input from the human supervisor.

This paper presents a description of our approach to the development of a machine vision/learning system and its learning methodology. Key algorithms are described in detail that highlights the system's unique nature and significant potential for image analysis. Results from a case study using Akamai in the analysis of indications of micro-calcifications are presented. Their significance for application of the system to other lesion types and other medical imaging applications are discussed. Performance considerations are discussed along with current and future directions for research and development.

2. CIPA – IMAGE PARTITIONING

Akamai implements some of the key functionality of the primate vision system [17,18], taking advantage of aspects that relate to efficient memory usage, learning from visual cues and image processing speed. A primate's retina has an arrangement of cones that is described by a hexagonal lattice [19]. The hexagonal architecture optimizes both information capture and error reduction by providing maximum receptor area with minimum inter-receptor space. Bees exploit this property to optimize the quantity of honey stored for the amount of wax used. This property, known as the honeycomb conjecture was not proven until recently by Peterson [20].

The concepts of space-variant sensing and the hexagonal lattice [19,20] were combined to form the underlying architecture of a new paradigm for artificial vision, named Spiral Architecture. The thrust of this paradigm is that it attempts to extract computational principles inherent in biological vision systems and implement them in digital technology. The mathematical structure of the Spiral Architecture is Lie Algebra and is described in [21].

Akamai takes advantage of the efficiencies of hexagonal architecture and multi-resolution processing by implementing CIPA, the "Constructive Image Partitioning Algorithm". Outlined below, and presented in detail in [22], the algorithm extracts descriptive attributes (equivalence classes) of the image by collecting together hexagonal pixels, which are contiguous and surrounded by a boundary consisting of pixels of similar intensities. Figure 1 displays a collection of seven hexagons of the lattice; where it can be observed that any three mutually adjacent hexagons form a Y-junction at their point of confluence.

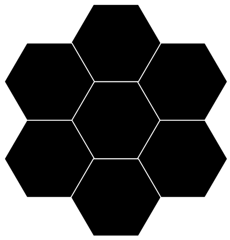

Figure 1. Hexagons arranged such that a center hexagon is adjacent to six other hexagons.

The algorithm provides a computational method to establish this boundary by tracing a path along the edges and thus between hexagons. The edge between two hexagons is called an "edgelet". The path is generated from an initial point by selecting the next path element (edgelet) from a choice of two at the Y-junction. The algorithm chooses the path by remaining between pixels with maximal intensity differential. The reader is referred to [22] for an explanation of why the method never involves an arbitrary decision in the choice of path elements.

The CIPA algorithm iteratively partitions the pixel data producing new equivalence classes at each repetition. The equivalence classes correspond to entities visible in the image by the human observer. The equivalence relation on the lattice is the property of connectedness; two adjacent hexagons are connected if their common edgelet is not part of a boundary.

At the first iteration, all hexagons are connected and thus form a single equivalence class. The path commences at the edgelet of a Y-junction separating the two pixels of maximum differential intensity. If both of the remaining edgelets of the Y-junction are not part of a boundary, then the edgelet associated with the larger of the two derivatives, is placed in a priority queue. The algorithm then repeatedly performs the following two steps:

- Remove an edgelet from the priority queue; if it is not part of a boundary, label it as a boundary and
- Place the edgelet corresponding to the larger of the Y-junction's two remaining edgelets, into the priority queue.

This boundary generating process terminates when the priority queue is empty. The closed boundary establishes a finer partitioning of the class by producing two new equivalence classes from the original.

A natural data structure to associate with the algorithm is a binary tree structure. Each node of the tree holds an equivalence class. The root of the tree represents the entire input image partitioned into a single equivalence class and thus possesses little visual information. The children of a node are the new equivalence classes that result from the boundary generated at the parent node. At the completion of each repetition of the algorithm, the collection of leaf nodes represents a partitioning of the image. Nodes at different levels of the tree represent views of segments of the image at different levels of resolution. Each leaf node of the completed tree represents an atomic visual entity.

3. MACHINE LEARNING IN AKAMAI

Mitchell defines a machine-learning algorithm as one that can learn from experience (observed examples) with respect to some class of tasks and a performance measure [12]. A learning algorithm can construct classifiers and/or hypotheses that represent and explain complex relationships in data.

Broadly, machine-learning schemes can be classified as either “unsupervised” or “supervised”. In unsupervised learning, no information is given to the learner about the data or the output and a set of programmed rules are followed to characterize, classify and cluster the output data. Supervised learning has (expert) knowledge about the data, its representation and characterization, and uses this *a priori* knowledge to classify data examples. *A priori* knowledge is accumulated through sets of training data, pre-classified into positive and negative examples of each concept to be learned.

Sufficient, quality examples need to be provided to ensure the learning algorithm can reach its required accuracy in terms of sensitivity (detection) and specificity (identification). Accounts have revealed that most individual learners are stronger in either sensitivity or specificity [14]. To ensure high sensitivity, a large range of representative, positive training examples may be required. Conversely, specificity is improved when an equal, or preferably larger number of negative training examples are supplied to the learner. These trends point to the requirement of a large amount of training data to ensure accurate induced classifiers.

Graphical data sets in medical imaging are a complex mixture of signals and noise, presenting a learning environment that is best suited to the supervised learning approach. Supervised learning methods can be classified as either rule-based, statistical or ensemble learning methods [13]. Rule-based methods (decision trees, version spaces, lazy learning, rule-based, etc) are ideal learners where classification is based upon discrete or categorical attributes. Statistical methods (naïve Bayesian networks, neural networks, support vector machines, etc) are ideal in situations where there are multiple dimensions to discern and where attributes are of a continuous nature. Each individual learning algorithm/method has its strengths and weaknesses.

Akamai has access to range of different learner modules that can be used to induce the required classifiers for mammogram analysis. Its current default learner is the decision tree and is currently being applied to detection and analysis of micro-calcification clusters. Other learners for making weighted or statistical decisions can also constructed using a Bayesian network and/or a neural network module. Future developments provide for the implementation of ensemble learners to better classify some of the more complex concepts in mammograms. Current work with the Akamai system is developing on three fronts and these are explained in greater detail in following sections of this paper.

4. GUIDING THE SUPERVISION PROCESS

In this section we describe an interactive, performance enhancing strategy (a process) that streamlines the acquisition of the training set from graphical data. In particular, a goal of this process is to maximize accuracy of classification and minimize the expenditure of resources in acquiring the training examples. One of the scarce resources in this process is the time taken by the human supervisor to acquire the training examples.

Our approach to achieving this goal is to initially build a classifier from special instances indicative aspects of the target concept provided directly by the supervisor. Then, progressively relax the supervisor’s responsibility for the identification of training instances as the power of the classifier improves. The technique described below embodies this strategy. Either the supervisor or Akamai can assume the responsibility for driving the process of acquiring training examples. In either case, as Akamai is presented with each training instance, it adds the instance to its training set and re-builds its classifier from the new set.

4.1 Supervisor-Driven Mode

In Supervisor-Driven mode, the supervisor takes full responsibility for the classification and order in which the artifacts are displayed. This responsibility can manifest in one of two sub-modes, Static and Dynamic.

4.1.1. Static Mode

The goal of “Static” mode is to generate a collection of key occurrences or views of the target concept. The collection should also contain examples of the target concept represented over the full range of resolutions employed by Akamai. The goal is achieved by having the supervisor interact with Akamai as described in the following process:

- The supervisor marks the boundary of a key instance of the target concept on an image presented on the GUI with the use of a mouse.
- The supervisor then instructs Akamai to foveate on the marked artifact.
- Akamai responds by searching through its internal representation of the image for the collection of pixels that most closely resembles the boundary of the marked artifact.
- Akamai then displays its artifact on the GUI so that the supervisor can visually compare Akamai's artifact with the marked up artifact.
- After a best match has been established, the supervisor classifies Akamai's artifact as one of four possible categories: 'Is`, 'Part`, 'Not` or 'Candidate`.
- The newly created training example is then added to the training set.

This Static mode is generally employed in the initial stages of the supervision process to generate positive training instances at high resolution and candidate instances at the lower resolutions.

4.1.2. Dynamic Mode

In "Dynamic" mode, the supervisor partially relinquishes to Akamai the responsibility to locate the training examples. The goal of Dynamic mode is to have Akamai learn candidate instances so that it can successfully determine when to foveate a candidate artifact. This implements a form of "reinforcement" learning and is achieved with Supervisor/Akamai interaction as described in the following process:

- Akamai traverses its internal representation of the image. The traversal corresponds to the sequence of artifacts as generated by CIPA.
- On display of each artifact, the supervisor classifies it appropriately. Each time the supervisor judges that the features of the current artifact represent a possible instance of the target concept but requires a view of the artifact at higher resolution, the classification of 'Candidate` is applied to the instance.
- At this point, Akamai pauses from the sub-tree traversal at the current resolution and attempts to locate the artifact at a higher resolution for the supervisor to classify.
- As each artifact is presented to the supervisor, Akamai makes a prediction with its latest updated classifier. Akamai compares its prediction with that of the supervisor's classification and keeps a running account of its error rate.
- This error rate is displayed on the GUI so that the supervisor can monitor Akamai's performance.

This mode is generally continued until such time as Akamai's error rate is sufficiently small; at which time, the supervisor changes the mode of supervision to move the learner/classifier on to the next most active and responsible role.

4.2. Akamai-Driven Mode

In the Akamai driven mode, the supervisor relinquishes further responsibility to Akamai for the learning process. Akamai drives the traversal of its internal representation from the current state of its classifier while the supervisor merely provides feedback to Akamai on its prediction of each artifact displayed. This mode has three sub-modes, "incremental", "next-positive" and "all-positive". Each of these sub-modes differs only in the amount of supervisor feedback provided to Akamai.

4.2.1. Incremental Mode

With operation in "incremental" mode the supervisor provides feedback on all artifacts that Akamai considers. The primary goal of the mode is to provide Akamai with feedback on its performance in identifying candidate instances and thus its ability to distinguish between the artifacts it should foveate and those that it should ignore. Supervisor feedback permits Akamai to recover from false positive predictions at lower resolutions, which would otherwise drive Akamai's traversal to higher resolutions unproductively. Incremental mode continues until such time as the supervisor deems that Akamai is identifying candidate artifacts sufficiently well; at which time the mode is switched to the more machine-centered Next-Positive mode.

4.2.2. Next-Positive Mode

In Next-Positive mode, Akamai requests feedback on each of the artifacts that it classifies as "positive". The goal of the feedback in this mode is to reduce Akamai's false positive error rate. This is achieved with Supervisor/Akamai interaction described as follows:

- Akamai traverses its internal representation of the image searching for candidate instances of the concept employing the current state of its classifier to distinguish between candidate/non-candidate artifacts.
- When it finds a candidate instance, it searches its internal representation at the next higher resolution for an artifact at the identified location in the image.
- In this process, if it finds an artifact that it classifies as a positive example of the concept, it displays it on the GUI and waits for supervisor feedback.

This mode continues until such time as the supervisor deems that Akamai is identifying instances of the concept at a sufficiently low error rate; at which time the mode is switched to All-Positive.

4.2.3. All-Positive Mode

In All-Positive mode the supervisor provides feedback only after Akamai displays all of the artifacts that it has classified as positive. The supervisor's goal is to correct all of Akamai's false positive and false negative classifications. To this end, upon Akamai's completion of its attempts to identify all occurrences of the target concept, the supervisor marks up artifacts on the GUI in a manner similar to the technique employed in Supervisor-Driven Static mode. When the supervisor completes this feedback process, a measure of Akamai's error rate is computed and displayed on the GUI. Akamai also has the opportunity to add the supervisor's feedback to its training set and re-build its classifier. This mode continues until the supervisor deems Akamai's overall performance is optimal. At this time Akamai's ability to identify and locate instances of the target concept is considered good enough to be employed without supervision.

5. CASE STUDY

Figure 2 displays a cropped mammogram containing micro-calcifications. The supervisor's task is to classify the nodes composing the tree structure of Akamai's internal representation as either positive or negative training examples of the target concept. In this case: “micro-calcification”.

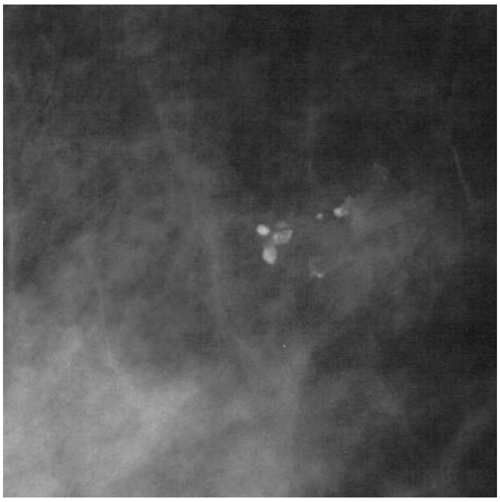

Figure 2. Cropped mammogram showing micro-calcifications

In this study, the CIPA tree structure for the mammogram contains approximately 1000 nodes. The number of nodes that correspond to micro-calcifications is only about 2 percent of the total. Initial use of *Supervisor-Driven Static* mode permitted these 20 nodes corresponding to positive instances of the target concept to be accessed directly and classified accordingly. The remaining 980 nodes were then explored in the modes with lower levels of human supervision.

In *Supervisor-Driven Dynamic* mode about 20 negative instances of the target concept were obtained to balance the number of positive and negative training instances. The supervisor then switched to *Akamai-Driven Incremental* mode with this initial classifier of micro-calcifications. Over the next 20 nodes, Akamai employed the classifier to correctly classify each of these negative instances. The supervisor then switched to *Akamai-Driven Next-Positive* mode to correct Akamai's classification of false-positive predictions. In this mode Akamai incorrectly moved to higher resolutions frequently. It was then concluded that more instances of ‘candidate’ were required and that these instances would be best obtained at *Supervisor-Driven Static* mode. In this case, the supervisor was not able to employ *Akamai-Driven All-Positive* mode due to the excessively high error rate in the mode below.

6. DISCUSSION

“Active” learning in Akamai is still only in early developmental stages but already demonstrates significant potential. While tentative results from the limited case study did not allow training to proceed to the lowest level of human supervision, it did demonstrate the feedback cycle that ensures learner accuracy.

Convergence in demonstrated learning and positive feedback is required before higher modes of machine driven learning are allowed. This ensures, progressively, that there are sufficient positive and negative examples to maintain both sensitivity and specificity at an acceptable level. This learning scheme has some similarity to elements of “reinforcement learning” [15,16] and seeks to minimise knowledge “noise” by seeking rule reinforcement, vision correction and corroboration of classification correctness.

Ostensibly, the same technique applied to classifying the micro-calcification concept can be applied to any lesion concepts in a similar way. What differs are the characterizing attributes of each concept and how much training data is required to learn the concept to an acceptable accuracy.

The need to make the input of training data more efficient is driving the development of a collaborative training paradigm with an effective collaborative user interface. Both the paradigm and interface, work in progress, are required to streamline the training data input and to make most effective use of trainer (supervisor/expert) time.

7. CONCLUSION

In this paper we have given an overview of the motivation for developing a computer-assisted method for detecting and diagnosing artifacts in medical images. In particular we have stressed its importance in application to the area of screening mammography and the need to improve the accuracy and timeliness of diagnosis of abnormal lesions.

Algorithms used in this machine vision/learning software are primarily biologically inspired. Sound justification is given for their development as a tool for human-supervised machine learning, particularly in the area of data embedded in complex images.

Machine-learning paradigms and strategies are discussed, in particular the “supervised” learning modes and the overhead that they exact in terms of supervision time. A progressive scale of supervision modes is described that concurrently ensures that sufficient training examples are entered to maintain standards of accuracy, and that the supervision process is executed in the most efficient manner. A case study is described that demonstrates the stages of supervision progression and the requirement for convergence towards consistent results before the machine-learner is accepted.

With the results of this preliminary case study we have demonstrated sufficient success to warrant further investigation of this new supervision and learning strategy.

8. REFERENCES

[1] Bray, F., McCarron, P., Parkin, D.M., “The changing global patterns of female breast cancer incidence and mortality”, *Breast Cancer Res*. v6 pp229-239, (2004), http://breast-cancer-research.com/content/6/6/229

[2] Mettlin C, “Global Breast Cancer mortality Statistics”, *CA Cancer J Cli*n, v49, pp135-137, (1999)

[3] National Breast Cancer Centre – Position Statement, “Early Detection Of Breast Cancer”, (2004) http://www.nbcc.org.au/media/early_detection.html

[4] D’Orsi, C.J., “Computer-Aided Detection: There Is No Free Lunch”, *Radiology*, v221, pp585-586, (2001)

[5] Zheng, B., et al., “Soft-Copy Mammographic Readings With Different Computer-Assisted Detection Cuing Environments: Preliminary Findings”, *Radiology*, v221, pp633-640, (2001)

[6] Carney, P.A., et al.,, “Individual And Combined Effects Of Age, Breast Density And Hormone Replacement Therapy Use On The Accuracy Of Screening Mammography”, *Annals Of Internal Medicine*, v138, n3, pp168-175, (2003)

[7] Barlow, W.E., et al., “Accuracy Of Screening Mammography Interpretation By Characteristics Of Radiologists”, *JNCI Journal of the National Cancer Institute*, v96, n24, pp1840-1850, (2004)

[8] Baines, C.J., Dayan, R., “A Tangled Web: Factors Likely To Affect The Efficacy Of Screening Mammography”, *JNCI Journal of the National Cancer Institute*, v91, n10, pp833-838, (1999)

[9] Taplin, S.H., Rutter, C.M., et al., “Accuracy Of Screening Mammography Using Single Versus Independent Double Interpretation”, *AJR American Journal Of Roentgenology*, v174, pp1257-1262, (2000)

[10] Keith, L.G., Oleszczuk, J.J., Laguens, M., “Are Mammography And Palpation Sufficient For Breast Cancer Screening: A Dissenting Opinion”, Lifeline BioTechnologies Inc, http://www.lbti.com/clinical2.asp

[11] Thurfjell, E., Thurfjell, M.G., Egge, E., Bjurstam N, “Sensitivity and specificity of computer-assisted breast cancer detection in mammography screening”, *Acta Radiologica*, v39, n4, pp384-388, (1998)

[12] Mitchell, T., “Machine Learning”, (1997), McGraw-Hill

[13] Dietterich, T.G., “Ensemble Methods in Machine Learning”, *First International Workshop on Multiple Classifier Systems*, New York: Springer Verlag, LNCS 1857, pp1-15, (2000)

[14] Last, M., Minkov, E., Maimon, O., “Improving Stability of Decision Trees", *International Journal of Pattern Recognitions and Applied Intelligence (IJPRAI)*, v16, n2, pp.145-159, (2002)

[15] Kaelbling, L.P., Littman, M.L., Moore, A.W., "Reinforcement Learning: A Survey", Journal of AI Research, Volume 4, pages 237-285, (1996)

[16] Moriarty, D.E., Schultz, A.C., Grefenstette, J.J., "Evolutionary Algorithms for Reinforcement Learning", *Journal of AI Research*, v11, pp241-276, (1999)

[17] Schwartz, E., “Computational Anatomy and Functional Architecture of Striate Cortex: A Spatial Mapping Approach to Perceptual Coding”, *Vision Research*, *20*, (1980).

[18] Bederson, B.,Walace, R., Schwartz, E., “A miniaturized space-variant active vision system: Cortex-1.”, *Machine Vision and applications*, v8, pp101-109. (1995)

[19] Williams, D.R., “Seeing through the photoreceptor mosiac”, *Trends in Neuro Sciences*, v9, n5 May, pp193-198, (1986)

[20] Peterson, I., “The honeycomb conjecture”, *Science News*, v156, n4, pp60-61, (1999)

[21] Sheridan, P., Hintz, T., Alexander, D., “Pseudo-Invariant Transformations on a Hexagonal Lattice, “*Image and Vision Computing”*, v18, n11, pp907-917 (2000)

[22] Sheridan, P., Hintz, T., Alexander, D., “Space-Variant Sensing For Robotic Vision”, *Proceedings of the 5th International Conference on Mechatronics and Machine Vision in Practice*, Nanging, pp185-190, (1998)