Precis of published work

Virginia R. de Sa
Department of Cognitive Science
University of California, San Diego
La Jolla, CA 92093-0515

de Sa, V.R., & Ballard, D.H. (1992). Top-down teaching enables task-relevant classification with competitive learning. In *IJCNN International Joint Conference on Neural Networks* (Vol. 3, pp. III-364—III-371).

de Sa, V.R., & Ballard, D.H. (1993). A Note on Learning Vector Quantization. In C.L. Giles, S.J. Hanson & J.D. Cowan (Eds.), *Advances in Neural Information Processing Systems 5*, (pp. 220—227). Morgan Kaufmann.

These papers studied aspects of learning algorithms that I went on to use in my thesis algorithms below. The first paper explores the benefits of adding extra "teaching" dimensions to the usually unsupervised algorithm of Competitive Learning.

In The second paper, I explore the differences between different versions of Kohonen's learning vector quantization (LVQ) algorithms, and show mathematically why the LVQ2.0 algorithm works better for separable distributions and the LVQ2.1 algorithm works better for overlapping distributions. In the course of the theoretical analysis, it is apparent that the arguments do not hold for the exact LVQ algorithms but would if certain modifications are made. I made these modifications and get improved performance. This paper follows logically from the first as Kohonen's LVQ algorithms are even better ways for adding supervision to Competitive style learning algorithms.

de Sa, V.R., & Ballard, D.H. (1993). Self-teaching through Correlated Input. In *Computation and Neural Systems 1992* (Chapter 66, pp. 437—441). Kluwer Academic Publishers.

de Sa, V.R. (1994). Minimizing Disagreement for Self-Supervised Classification. In M.C. Mozer, P. Smolensky, D.S. Touretzky & J.L. Elman (Eds.), *Proceedings of the 1993 Connectionist Models Summer School* (pp. 300—307).Erlbaum Associates.

de Sa, V.R. (1994). Learning Classification with Unlabeled Data. In J.D. Cowan, G. Tesauro, and J. Alspector (Eds.), *Advances in Neural Information Processing Systems 6* (pp. 112—119). Morgan Kaufmann.

de Sa, V.R. , & Ballard, D. (1997). Perceptual Learning from Cross-Modal Feedback. In R. L. Goldstone, P. G. Schyns, & D. L. Medin (Eds.) Psychology of Learning and Motivation, Vol 36. (pp 309–351) San Diego, CA: Academic Press.

de Sa, V.R., & Ballard, D.H. (1998). Category Learning through Multi-Modality Sensing. *Neural Computation* 10(5), 1097–1117.

de Sa, V. (1999). Combining Uni-Modal Classifiers to Improve Learning. In H. Ritter, H. Cruse, & J. Dean (Eds.) Prerational Intelligence: Adaptive Behavior and Intelligent Systems without Symbols and Logic, Vol 2. (pp 709-723) Dordrecht, The Netherlands: Kluwer Academic Publishers.

This series of papers formed the core of my dissertation research. The papers develop a fully unsupervised (other than picking a number of categories) algorithm that performs comparably to supervised algorithms on categorization tasks. The idea behind the papers is that when an object is experienced, sensations in different modalities are correlated and these correlations can be usefully used to help separate classes within each individual modality. The different modalities attempt to minimize the disagreements in their output classifications (subject to a pressure to keep some patterns in each class). I developed a way to use this information with an algorithm that uses ideas from the LVQ algorithms above.

The basic idea is presented in the first paper. The equations are developed in the second paper and performance results are given in the third paper. The fourth and sixth articles are invited book chapters where the material is presented for specialized audiences and the fifth paper presents the algorithm using biologically plausible computations.

These papers are examples of using ideas from the architecture of the brain (separation of different modalities, feedback connections, Hebbian learning) to develop a new machine learning algorithm that fills a new niche of unsupervised learning. In addition I am now using the model to understand why sensory modalities may be separated the way they are (first paper below). This paper shows that for this form of self-supervised learning it is very important to keep related dimensions on the same side of the network. It is better to throw away valuable auditory information than to add it to the visual side of the network. This is because it is important for the two sides of the network to be classifying based on very different features.

My latest paper in this area (second paper below) develops a spectral clustering version of my thesis algorithm. In the spectral clustering domain it is easy to see the relationship between the idea of minimizing disagreement between two sources and just clustering in the combined joint space. The different benefits of each algorithm are explored and I show that the minimizing-disagreement spectral clustering algorithm works better than other options when some patterns have only one view available. I have since expanded the spectral clustering algorithm to work with multiple views.

This work most strongly demonstrates the interplay between machine and brain learning.

de Sa, V.R. (2004). Sensory Modality Segregation. To appear in S. Thrun, L. Saul, and B. Schoelkopf (Eds.), *Advances in Neural Information Processing Systems 16.* (pp. 913–920). MIT Press.

de Sa, V.R. (2005). Spectral Clustering with Two Views. ICML (International Conference on Machine Learning) Workshop on Learning with Multiple Views. Bonn, Germany.

Caruana, R., & de Sa, V.R. (1997) Promoting Poor Features to Supervisors: Some Inputs Work Better as Outputs. In M.C. Mozer, M.I. Jordan and T.P. Petsche (Eds.), *Advances in Neural Information Processing Systems 9.* (pp. 389–395). MIT Press.

Caruana, R., & de Sa, V.R. (1998). Using Feature Selection to Find Inputs that Work Better as Outputs. In the proceedings of the 8th International Conference on Artificial Neural Networks (ICANN 98), Skövde, Sweden. (pp. 299–304). Springer-Verlag London.

Caruana, R., & de Sa, V.R. (2003). Benefitting from the Variables that Variable Selection Discards. *Journal of Machine Learning Research* 3(Mar):1245–1264, 2003.

de Sa, V.R. , Morgan, R.J., & Caruana, R. Using Reverse feature selection to find features that work better as extra outputs than as inputs. Unpublished manuscript.

In this series of papers, we establish the usefulness of what we call output feature selection.

People have found that the generalization performance of the learning systems can be improved by removing dimensions of the input patterns during training (and testing). This is known as dimensionality reduction or feature selection). What we show in the first paper above is that there are easily understood cases where rather than discard the unused dimensions, we can get overall performance improvements if we use those extra dimensions as extra outputs. The way extra outputs can help is that they also back-propagate error to the hidden layer which will influence the mapping from the input to hidden layer.

In the second paper above, we demonstrate this for a real problem of DNA splice junction recognition. Improving performance on this problem is notable as the problem has been in a popular machine learning database for a long time. It was also important to demonstrate that the theoretical arguments we made in the paper above actually hold in a real problem. The third paper is in the JMLR special issue on feature selection and combines the findings of both papers as well as reporting results on an important real problem of pneumonia risk assessment where we got similar performance improvements.

The last manuscript, we (including Robert Morgan— one of my undergraduate students who worked on this project as a 199 and summer student) present a principled approach to selecting the features to be used for extra output features. In the manuscript we show improved performance on the DNA splice junction recognition when we follow the algorithm, than when we used the naive approach from the earlier paper.

These papers are related to the series of cross-modal papers above because they both address the issue of how information should be optimally presented to a network for learning. In the multi-modality work, I find a big advantage to processing the auditory inputs on one side of the network and the visual inputs in another. In these papers, we are addressing the same question, in the supervised domain, and in cases where the divisions are not so obvious and must be discovered.

Saygin, A.P., Driver, J., & de Sa, V.R. In the footsteps of biological motion and multisensory perception: Judgements of audio-visual temporal relations are enhanced for upright walkers under review at *Psychological Science*.

In my thesis work, I showed that cross-modal associations could be helpful for developing sensory systems. In this paper we show a new example of learned cross-modal associations affecting human performance. We show that when subjects are doing an auditory-visual task that involves natural temporal correlations, they perform much better than when the natural relationships are perturbed. In particular we show that subjects are better able to match the frequency of visual and auditory footsteps when a point-light walker is upright, than when it is inverted. This improvement is only there when the sounds are in the correct relationship to be footprints than when they are not. This paper shows that learned cross-modal associations can affect everyday performance in related tasks.

Trottier, L.G., & de Sa, V.R. (2007). A Multimodal Paradigm for Investigating the Perisaccadic Temporal Inversion Effect in Vision. to appear in the *Proceedings of the 29th Annual Meeting of the Cognitive Science Society*.

This paper marks the beginning of our study of the how visual input is integrated between saccades. In this paper we attempt to shed more light on the temporal inversion illusion discovered by Concetta-Morrone and colleagues (Nature Neuroscience 2005) — when two flashes occur within 40 msecs of each other and approximately 30 to 70 msecs prior to a saccade the perceived order of the flashes reverses (the one presented first is perceived last). We show that auditory clicks are not subject to this inversion and then use an auditory click as a relative time stamp to examine what happens to an individual visual flash as a function of presentation time relative to saccade onset. We showed that the differences in perceived time of flashes earlier in the saccadic preparation window (more than 50 msecs before a saccade) and those later in the window (0 to 50 msecs prior to a saccade) can explain the temporal inversion illusion. We are currently working on a computational model that model that links this result to the findings of receptive field remapping (Duhamel, Colby and Goldberg 1992) and saccadic supression.

---

de Sa, V.R., & Hinton, G.E. (1998). Cascaded Redundancy Reduction. *Network: Computation in Neural Systems* 9(1), 73–84.

This paper develops a new unsupervised clustering algorithm that clusters patterns through creating a generative model (through feedback connections) that can explain the patterns. It has a major advantage over many other algorithms in that it incrementally adds clusters and has a well-defined cost function to determine when to stop dividing clusters.

de Sa, V.R., deCharms, R.C., & Merzenich, M.M. (1998) Using Helmholtz Machines to analyze multi-channel neuronal recordings. In M.I. Jordan, M.J. Kearns, and S.A. Solla (Eds.), *Advances in Neural Information Processing Systems 10*. (pp. 131–137). MIT Press.

This paper extends the algorithm above to address the important question of whether neurons code independently. We used the algorithm above to look for clusters in the firing patterns across multiply recorded neurons. We demonstrate the algorithm on synthetic data to show that it can find patterns when they exist (and under what conditions) and more importantly that it doesn't find patterns that weren't inserted in to the synthetic data. We then applied the algorithm to recordings from new world monkeys and found patterns in the data. The patterns corresponded to correlated rate modulations between some subset of the channels. This is another example of using machine learning algorithms to get more benefit from neuroscience experiments.

Yu, H-H., & de Sa, V.R. (2004). Nonlinear reverse-correlation with synthesized naturalistic noise. *Neurocomputing* 58-60:909–913.

This paper similarly uses sophisticated computational analysis to gain more information from neuroscience experiments. In this paper, Hsin-Hao develops the idea that "naturalistic" noise can be used to automatically determine receptive field properties in early visual neurons. Natural scenes are too complicated for analysis and simple white noise stimuli do not adequately stimulate neurons beyond the earliest cortical neurons. Equations are derived to allow calculations of kernels from the responses to the naturalistic noise stimulus. It is hoped that this will greatly improve our ability to automatically determine receptive field properties.

Hammon, P.S. & de Sa, V.R.(2007). Pre-processing and meta-classification for brain-computer interfaces. *IEEE Transactions on Biomedical Engineering* 54(3): 518–525. Digital Object Identifier: 10.1109/TBME.2006.888833

Hammon, P.S., Pineda, J.A., & de Sa, V.R. (2006). Viewing motion animations during motor imagery: effects on motor imagery. In G.R. Mueller-Putz, C. Brunner, R. Leeb, R. Scherer, . Schloegl, S. Wriessneggger, and G. Pfurtscheller, Proceedings of the 3rd International Brain-Computer Interface Workshop and Training Course 2006, pages 62-63, 2006.

Hammon, P.S., Makeig, S., Poizner, H., Todorov, E., & de Sa, V.R. (2008) Predicting Reaching Targets from Human EEG. to appear in IEEE Signal Processing Magazine (special issue on Brain-Computer Interfaces Jan 2008).

These three papers are a new twist in the area of applying machine learning to better understand neural recordings. We have started working in Brain-Computer Interfaces using non-invasive EEG recordings. Our plan is to have the user imagine natural movements and to provide natural-like visual feedback. We believe that the mapping from natural (imagined) movements to realistic feedback of the intended movement will be easier and more intuitive to learn and use than the somewhat arbitrary pairing of mental task and feedback currently employed in most BCI systems.

The first paper introduces our automated method for creating an accurate BCI. We create many sensible features based on known properties of EEG signals and use efficient and careful cross-validation we determine which set of features performs best. We then combine many good-performing feature vectors using meta-classifiers to obtain our final classifier. This strategy won a second-place finish in the 3rd International Brain-Computer Interface competition. The paper also analyzes and discusses the relative benefits of different signal processing and other pre-processing benefits for three different EEG/ECoG datasets.

The purpose of the work in the second paper was to determine whether our general strategy for improving Brain-Computer Interfaces seemed feasible. It is known that the EEG reveals a decrease in power in the mu frequency band (mu desynchronization) during movement. Mu desynchronization can also be seen during imagined movement and when watching movies of movement. In our paper we show that these two effects can add to give a stronger mu desynchronization than imagination alone. This is important because it means that once we start to detect a mu desynchronization, we can further increase the signal by providing visual feedback of the detected motion.

In the third paper we apply our EEG feature classification techniques to the task of predicting movement endpoints. We show that EEG recordings contain enough information to determine reaching targets both during (with appropriate artifact removal) and prior to planned arm reaches.

Sullivan, T.J., & de Sa, V.R. (2004). A Temporal Trace and SOM-based Model of Complex Cell Development. *Neurocomputing* 58-60:827–833.

This paper tests the idea that a *map* of complex cells could develop from a temporal trace rule (local memory) and retinal waves (as have been observed in developing cats). An earlier paper by Foldiak developed four complex cells from four types of simple cells, waves of only four orientations (and two directions) and a temporal trace rule. We use tuned simple cells, present waves of random orientation, and develop a topographic map of complex cells.

Sullivan, T.J., & de Sa, V.R.. Unsupervised Learning of Complex Cell Behavior. in preparation.

This paper extends the work above by using more realistic neural models. The learned model neurons are phase-invariant, tuned for orientation, tuned for spatial frequency and form an orderly map.

Sullivan, T.J., & de Sa, V.R. (2006). A Model of Surround Suppression through Cortical Feedback. Neural Networks 19(5), 564–572.

This paper presents a model that explains the seemingly contradictory findings that direct top-down feedback connections (e.g. from visual area V2 to visual area V1) are almost exclusively excitatory (from excitatory neurons to excitatory neurons) and yet in vivo, the finding has been that feedback connections tend to slightly increase excitation to a stimulus in the center of the receptive field but greatly inhibit responses to stimuli in the surround of a neuron. The model replicates findings from different physiological experiments.

Sullivan, T.J. & de Sa, V.R.(2006). Homeostatic synaptic scaling in self-organizing maps. *Neural Networks* , Volume 19, Issues 6-7, Pages 734-743 (July-August 2006) Advances in Self Organising Maps - WSOM05 Edited by Marie Cottrell and Michel Verleysen.

Sullivan, T.J. & de Sa, V.R.(2006). A self-organizing map with homeostatic synaptic scaling. *Neurocomputing* Volume 69, Issues 10-12, June 2006, Pages 1183-1186 .

Sullivan, T.J., & de Sa, V.R. Sleeping our way to stable learning: a theoretical study. under review at *Neural Computation*.

Recently it has been found that neurons tend to regulate their synapse strength depending on their activity. The above three papers investigate the computational properties of this biologically found rule. The first paper paper shows that the homeostatic plasticity rule can be used in place of the standard computationally convenient but otherwise unmotivated weight normalization method. We show that the rule is able to maintain consistent output firing rates even with large-scale proliferation and die-off. The second paper above expands the usefulness of this algorithm for simulations by deriving an equation that shows how to scale the learning rate as a function of the architecture parameters. The last paper shows that performing homeostatic synaptic scaling during sleep, leads to stable learning and is consistent with several sleep research findings, such as decreased slow wave activity throughout the night, increased need for sleep during periods of increased learning, and instability with insufficient sleep. We also show how the rule is mathematically related to the standard weight normalization rule.

Robinson, A.E., Hammon, P.S., & de Sa, V.R. (2007). Explaining brightness illusions using spatial filtering and local response normalization. *Vision Research* 47(12): 1631–1644.

The previous papers took the tack of starting with biologically plausible networks and computation rules and explaining observed physiological findings. In this paper we start from a very successful model (the ODOG model of Blakeslee and McCourt) and make it more biologically plausible. The ODOG model has accounted for numerous brightness illusions. We show that by making it more biologically plausible (by restricting computations to be more local both in space and frequency), we can actually account for more illusions than the original ODOG model.

---

de Sa, V.R., & Stryker, M.P. (2001) Cortical feedback projections and plasticity in a V1/V2 mouse slice. Abstract in the Society for Neuroscience 31st Annual Meeting, San Diego, CA. 2001.

This abstract presents results from the development of a V1/V2 mouse slice (from mice 28-34 days old) with intact connectivity between V1 and V2. We hypothesized that the feedback connections might act as teaching inputs as in the self-supervised model in my thesis work. We found a slight trend in that direction but nothing significant. We did however find that the feedback connections did not modify under protocols that potentiated the forward and local horizontal connections. Even when the horizontal input was reduced and delayed to look to have the strength and timing of feedback input, there was no consistent plasticity. This is consistent with modeling predictions that the feedback connections should not change when the forward ones do. It remains to be seen if a protocol for modifying the feedback connections can be found.

de Sa, V.R., & MacKay, D.J.C. (2001). Model fitting as an aid to bridge balancing. *Neurocomputing* (special issue devoted to Proceedings of the CNS 2000 meeting) Vol38-40, 1651–1656.

This paper was written to be an automated solution to the problem of ”balancing the bridge” in neurophysiological recording. This is a manipulation that is required in order to compensate for electrode resistance, so that the electrical potential across the cell membrane can be accurately measured. It is important to know the actual membrane potential because slight differences in the potential can make large differences in the size of the synaptic events and other measurements you make. In the paper, we use the maximum-likelihood approach to fit the voltage trace obtained by the usual application of a step current change.

This paper is another example of using machine-learning techniques to improve our collection of neural data.

McRae, K., de Sa, V.R., & Seidenberg, M.S. (1993). Modeling Property Intercorrelations in Conceptual Memory. In *Proceedings of the 15th Annual Meeting of the Cognitive Science Society* (pp. 729—734).

McRae, K., de Sa, V.R., & Seidenberg, M.S. (1997). On the Nature and Scope of Featural Representations of Word Meaning. *Journal of Experimental Psychology: General*, 126(2), 99–130.

McRae, K., Cree, G.S., Westmacott, R., & de Sa, V.R. (1999) Further Evidence for Feature Correlations in Semantic Memory. *Canadian Journal of Experimental Psychology* Special Issue on Word Recognition. 53(4), 360–373.

These papers were the result of a collaboration started at the University of Rochester with

Ken McRae. We wanted to explain the puzzling difference between the results Ken had found for the way people process words that correspond to natural objects vs words that correspond to man-made objects. I constructed a model that explained all of his behavioral differences (reaction time, semantic priming, and property verification latency). The model also showed a few extra features such as filling in properties that people just hadn't mentioned (such as: BUDGIE-has wings, TROUSERS-worn by women, MOUSE-has four legs)as well as making some interesting mistakes (such as thinking that a JET has feathers).

Zheng, C.L., de Sa, V.R., Gribskov, M., & Nair, T.M. (2003). On Selecting Features from Splice Junctions: An Analysis Using Information Theoretic and Machine Learning Approaches. In M. Gribskov, M. Kanehisa, S. Miyano, and T. Takagi (Eds.), *Genome Informatics* Vol. 14, Universal Academy Press, Inc.

This paper was the first result of a collaboration with biologists at the San Diego Supercomputing Center. In this paper we compare the results of T.M. Nair, C. Zheng and M. Gribskov using the neural network calliper technique developed by Nair with my results of looking at the information gain measure (previously used by many including Caruana & de Sa) applied to the problem of finding splice junctions in DNA sequences. The neural network calliper technique is noisier than the information gain measure but is easily able to consider the role of pairs and larger groups of nucleotides. This was one of 25 papers chosen to be a full peer-reviewed paper.
