# Knowledge-based Cascade-correlation: A Review

**François Rivest**
Département d'Informatique et
de Recherche Opérationnelle
Université de Montréal
CP 6128 succ. Centre Ville
Montréal, QC, Canada
H3C 3J7
*francois.rivest@mail.mcgill.ca*

**Thomas R. Shultz***
Department of Psychology and
School of Computer Science
McGill University
1205 Dr.Penfield Avenue
Montréal, QC, Canada
H3A 1B1
*thomas.shultz@mcgill.ca*

## Abstract

KBCC is an extension of the cascade-correlation algorithm that treats functions encapsulating prior knowledge as black-boxes which, like simple sigmoidal neurons, can be recruited in the network topology. KBCC has been studied on artificial and real tasks and it has successfully reused various kinds of knowledge. This paper surveys the work on KBCC, from old published data to the latest new results. It also describes KBCC's position in the transfer of knowledge. Likely future research is forecast.

## 1 Introduction

In 2000, a new algorithm for transfer of knowledge was presented at IJCNN [1] and ICML [2] called knowledge-based cascade-correlation (KBCC). This algorithm is an extension of the cascade-correlation algorithm and can insert prior knowledge directly into its topology as needed. Here, we review KBCC, its various applications, and its position as a transfer-of-knowledge algorithm.

## 2 The KBCC Algorithm

KBCC is a natural generalization of cascade-correlation (CC) [3]. CC is a neural network learning algorithm similar to backpropagation (BP), but instead of starting with a complete network topology, it begins with only input to output weights and adds hidden nodes during training as needed. Whereas CC is only able to recruit sigmoidal hidden units, KBCC is able to recruit whole trained networks into its architecture. Figure 1 shows an example of a KBCC network. Like CC, a hidden node can be installed beyond previous nodes, thus being fed by those older nodes. Any differentiable multivariate vector-valued function can be recruited. Throughout this paper, we call the network to train the *target* network and prior knowledge *source* networks.

*Correspondence should be send to Thomas Shultz *thomas.shultz@mcgill.ca.*

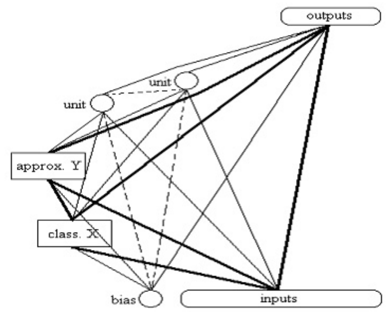


Figure 1: A KBCC network with four hidden units: a source classifier X, a source approximator Y, and two single sigmoid units. (Modified from [a])

Initially, a KBCC network has only input and output nodes fully connected using small random weights. Like CC, it alternates between output and input phases. In output phases, only the weights connecting to the output units are trained to minimize the sum squared error (equation 1). In input phase, weights feeding the pool of candidates are randomly initialized and trained to maximize the candidate's correlation with target-network output error (equation 2). Optimization is typically done using QuickProp. When the optimization process stagnates, or after a fixed number of epochs, training shifts from one phase to the other until final criterion is reached.

$$E = \sum_o \sum_p \left(V_{o,p} - T_{o,p}\right)^2 \tag{1}$$

where $V_{o,p}$ is the network output value for output $o$ of pattern $p$ and $T_{o,p}$ is the target (or desired) value for output $o$ of pattern $p$.

$$G_c = \frac{\left\| Cov\left(V_c, \left(V - T\right)\right)\right\|^F}{E} \tag{2}$$

where $\|\ \|^F$ is the Frobenius norm of the covariance matrix between the candidate output patterns $V_c$ of candidate $c$ and the error patterns $V$-$T$ of the target network and where $E$ is as in equation 1.

A detailed description of the algorithm with all the default parameters values can be found in [4][5]. Candidates can also be allowed to compete to be either on the latest top-most layer or on a new one using the sibling-descendant method [6]. Pruning methods can also be used on KBCC networks [7].

# 3 Survey of Experiment and Results

Research has focused on whether or not KBCC is able to use prior knowledge in learning, whether such knowledge can accelerate learning, and whether knowledge could compensate for lack of data or noisy data.

## 3.1 Using Prior Knowledge of Various Types

Almost every KBCC network we studied did recruit source networks if they were available. Our most systematic studies of KBCC [4] ([1][2][8]) used dichotomous classifications of patterns in a bounded bi-dimensional input space. Points inside a region of the space of a particular shape are *true*, while the others are *false*. KBCC recruited sources having a *true* cluster at a different location in the input space (translation), of a different size (scaling), or having a different orientation (rotation). It can combine simple sources to learn a more complex shape and reversely, can extract knowledge from complex sources to learn a simpler shape.

Combining simple sources into a complex solution is called *compositionality*. KBCC was able to learn a higher *parity* problem and a more differentiated *chessboard* problem, knowing simpler version of them [9]. KBCC was also able to learn how to test for prime numbers using knowledge of factors [in progress] and learn CAR-CDR compositions in Lisp [unpublished].

Sources can also be recurrent networks (in CAR-CDR). KBANN has also been used to convert symbolic rules into a differentiable network form recruitable by KBCC [7]. A main advantage over simply retraining the resulting KBANN network is that KBCC can select only the most useful rules for the given task [10][7]. KBCC has also been demonstrated successful in real world problems such as vowel recognition [5] and DNA splice-junction determination [10][7].

### 3.2 Accelerating Learning and Compensating for Lack of Data

KBCC was shown to accelerate learning of the foregoing tasks (see Figure 2). Compensation for lack of data is currently under study for the dichotomous cluster task. Figure 3 shows preliminary results for exact knowledge, irrelevant knowledge and absence of knowledge on a rectangular cluster as the training set size is reduced. One can see that as the amount of data decreases, there is better learning when very relevant knowledge is used as opposed to irrelevant or no knowledge.

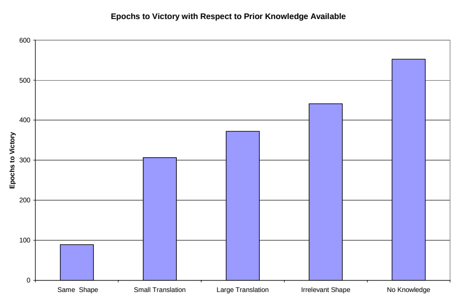


Figure 2: Knowledge speed-up.

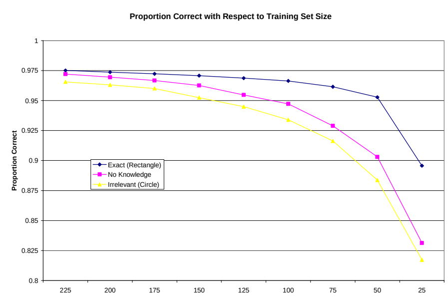


Figure 3: Knowledge-biased learning accuracy.

## 4 KBCC Position in the Knowledge Transfer Arena

Previous reviews of transfer of knowledge have focused on categorizing the type of transfer (representational or functional, direct or indirect, etc) [11]. In such a scheme, KBCC should be considered a direct literal representational transfer algorithm, because the original representation is copied unchanged into the new network topology. KBCC is less constrained than other representational transfer algorithms because only the derivative of the source knowledge is needed in KBCC. Even this constraint could be relaxed if the optimization method was not gradient based.

Unlike many knowledge-transfer algorithms, KBCC is not constrained to require the same input and output representations or the same internal topology across source and target networks. KBCC may be the only neural algorithm able to create cascaded compositions of multiple source networks.

Knowledge-transfer algorithms must select a relevant source, build and evaluate a map between source and target and update the knowledge pool to include the resulting new knowledge [12]. So far, the KBCC pool of source candidates contains knowledge the experimenter places there. Then KBCC by itself finds maps, evaluates and selects sources, and uses the best mapped sources to learn the target task. Perhaps KBCC could update the source pool by automatically storing the target network. At this point, selection is not fully automatic, because KBCC is unlikely to scale well under a large number of sources.

# 5 Conclusion and Future works

In short, KBCC was shown to be able to select and map a variety of prior knowledge in new learning. KBCC makes complex combinations of prior knowledge, and uses knowledge to accelerate learning and improve accuracy in cases of impoverished training data. KBCC was also shown to work on real world problems. Although some of these abilities still require deeper evaluation, an important next step is to look at life-long learning, or how to deal with a continually increasing bank of knowledge.

## Acknowledgments

This work was supported by grants from the Natural Sciences and Engineering Research Council to TRS. We are grateful to JP Thivierge for rapid answers to questions about his work.

## References

[1] Shultz, T.R. & Rivest, F. (2000) Knowledge-based Cascade-correlation, *IEEE-INNS-ENNS International Joint Conference on Neural Network 2000*, pp. V641-V646.

[2] Shultz, T.R. & Rivest, F. (2000) Knowledge-based Cascade-correlation: An Algorithm for Using Knowledge to Speed Learning. *Proceedings of the Seventeenth International Conference on Machine Learning*, pp. 871-878. San Francisco, CA: Morgan Kaufmann.

[3] Fahlman, S.E. & Lebiere, C. (1990) The Cascade-correlation Learning Architecture. In *Advances in Neural Information Processing Systems* **2**, pp. 524-532. Los Altos, CA: Morgan Kaufmann.

[4] Shultz, T.R. & Rivest, F. (2001) Knowledge-based Cascade-correlation: Using Knowledge to Speed Learning, *Connection Science* **13**:1-30.

[5] Rivest, F. & Shultz, T.R. (2002) Application of Knowledge-based Cascade-correlation to Vowel Recognition. *IEEE International Joint Conference on Neural Network 2002*, pp. 53-58.

[6] Baluja, S., & Fahlman, S.E. (1994). Reducing Network Depth in the Cascade-correlation Learning Architecture. *Technical Report* CMU-CS-94-209, School of Computer Science, Carnegie Mellon University.

[7] Thivierge, J.P., Dandurand, F., & Shultz, T.R. (2004). Transferring domain rules in a constructive network: Introducing RBCC. *IEEE International Joint Conference on Neural Networks 2004*, pp. 1403-1409.

[8] Shultz, T. R., & Rivest, F. (2003). Knowledge-based cascade-correlation: Varying the size and shape of relevant prior knowledge. In H. Yanai, A. Okada, K. Shigemasu, Y. Kano, & J. J. Meulman (Eds.), *New developments in psychometrics*, pp. 631-638. Tokyo: Springer-Verlag.

[9] Rivest, F., & Shultz, T.R. (2004) Compositionality in a Knowledge-based Constructive Learner. *Papers from the 2004 AAAI Symposium*, *Technical Report* FS-04-03, pp. 54-58. AAAI Press.

[10] Thivierge, J.-P., & Shultz, T.R. (2002). Finding relevant knowledge: KBCC applied to splice-junction determination. *IEEE International Joint Conference on Neural Networks 2002*. pp. 1401-1405.

[11] Pratt, L.Y. & Jennings, B. (1996). A Survey of Transfer Between Connectionist Networks. *Connection Science* **8**(2):163-184.

[12] Forbus, K., & Gentner, D. (1989). Structural evaluation of analogies: What counts? *Proceedings of the Cognitive Science Society*. Mahwah, NJ: Lawrence Erlbaum Associates.