Proceedings of the International Conference on Cognitive Modeling (2019).

Evolutionary Optimization of Neural-Network Models of Human Behavior

Uli Grasemann **Risto Miikkulainen**
Department of Computer Science
The University of Texas at Austin, Austin, TX 78712, USA

Claudia Peñaloza **Maria Dekhtyar** **Swathi Kiran**
Aphasia Research Laboratory, Department of Speech, Language, and Hearing Sciences
Boston University, Boston, MA 02215, USA

Abstract

Neural network models are essential tools in understanding how behavior arises from information processing in the brain. Recent advances in computing power and neural network algorithms have made more complex models possible, increasing their explanatory power. However, it is difficult to make such models work: they have many configuration parameters that have to be set right for the model to work properly. Consequently, automated methods are needed to optimize them. This paper proposed an evolutionary approach to this problem. An Age-Layered Evolutionary Algorithm is introduced and evaluated by fitting training parameters for BiLex, a self-organizing map model of lexical access in bilinguals. The resulting configurations are highly optimized and able to generalize to previously unseen human data, showing that evolutionary optimization of complex models has the potential to play an integral role in cognitive modeling in the future.

Keywords: Neural Networks; Cognitive Modeling; Evolutionary Algorithms

Introduction

Over the last few decades, connectionist neural networks have become an essential tool to characterize and investigate human cognition. Models based on such networks are usually not intended as physiologically accurate simulations of biological neurons and their interactions; nevertheless, they exhibit many characteristics of information processing in biological systems, including robustness to damage and input errors, and the ability to learn and generalize. This property of brain-like information processing on an abstract level is the main advantage of neural network-based models, enabling them to capture many aspects of high-level cognition while relying on mechanisms that are plausible analogs of the underlying neural substrate.

Recent progress in computing technology, such as GPU computing and software frameworks that rely on it, like Theano and TensorFlow (Abadi et al., 2015; Theano Development Team, 2016), have dramatically increased the performance and complexity of achievable models. At the same time, advances in neural network algorithms and architecture like deep learning and reservoir methods (Schmidhuber, 2014; Maass, Natschlager, & Markram, 2002) have made use of these capabilities, and thus the scale and performance of neural network applications have increased in equal measure.

Together these advances can significantly improve cognitive neural network models. Most importantly, rather than simulating behavior on an abstract and qualitative level, sufficiently large and complex networks can now be built so that clinical and psychometric tests can be modeled directly and quantitatively. Furthermore, rather than demonstrating that a certain kind of function of behavior can plausibly occur in a model, modern architectures can be used to investigate the link between environmental factors on the one hand, and the resulting individual differences on the other.

Building this new brand of models presents new and unique challenges. Most importantly, their ability to capture individual differences in behavioral data makes them sensitive to a large set of interdependent parameters governing e.g. module sizes, extent and intensity of training and pre-training, and input/output behavior of different classes of artificial neurons. In contrast to typical models in the past, fitting a model's many parameters manually in order to account for behavioral data is no longer feasible.

Another significant challenge is that the amount of individual human data available is often limited. Since the required amount increases with larger parameter spaces, and since quantitative measures need to be elicited for both target behavior and any individual differences of interest, acquiring the data necessary for accurate parameter fitting becomes prohibitively difficult.

Third, for an interdependent set of parameters that influence the behavior of the model in a non-linear way, fully evaluating a given set of model parameters involves training a complete model for each human participant. The resulting goodness-of-fit measure provides no gradient w.r.t. the parameter set. Therefore, the standard gradient based methods of metalearning cannot be used to optimize these models.

This paper proposes an evolutionary approach to these issues. The goal is to make parameter fitting of complex neural network models to limited human data workable in practice. In order to limit the cost of evaluation, the proposed EA uses a variant of the previously introduced Age-Layering technique (Shahrzad, Hodjat, & Miikkulainen, 2016), which aims to focus detailed evaluations on the most promising candidates.

The approach is evaluated in optimizing parameters for BiLex, a neural network model of the bilingual lexicon (Anon et al., 2016). BiLex simulates tests used in clinical practice, and captures the complex interactions between exposure to different languages and the resulting individual differences in bilingual lexical access. It is a complex model of individual subjects, for which little training data is available. It is therefore an appropriate test case for the proposed approach.

The next section gives an overview of bilingualism and the BiLex model. Using BiLex as a working example, the following sections then introduce and evaluate the proposed model fitting method, and discuss the results.

Bilingualism and the BiLex Model

The mental lexicon, i.e. the storage of word forms and their associated meanings, is a central component of language processing. The lexicon of bilinguals is considerably more complex than that of monolinguals, and the ways in which multiple language representations can develop, coexist, and interact are incompletely understood.

Given that the majority of the world's population is bilingual or multilingual (Bhatia & Ritchie, 2005), extending existing modeling approaches to improve our understanding of these additional complexities is of considerable practicle importance, and computational models of the bilingual lexicon could contribute to novel approaches in bilingual research, education, and clinical practice.

Current theoretical models of the bilingual lexicon generally agree that bilingual individuals have a shared semantic (or conceptual) system, and that there are separate lexical representations of the two languages (L1 and L2). However, the models differ on how L1 and L2 interact with the semantic system and with each other. The most recent model is the Revised Hiearchical Model, proposed by Kroll & Stewart (Kroll & Stewart, 1994). It assumes connections of varying strength between all three components, depending on relative language dominance.

The physiological structure and location of the lexicon in the brain are still open to some debate, but converging evidence from imaging, psycholinguistic, computational, and lesion studies suggests that the lexicon is laid out as one or several topographic maps, where concepts are organized according to some measure of similarity (Caramazza, Hillis, Leek, & Miozzo, 1994; Spitzer et al., 1998).

Self-organizing maps (SOMs; Kohonen, 2001) are neural networks that model such topographical structures, and are therefore a natural tool to build simulations of the lexicon. SOM models have been developed to understand e.g. how ambiguity is processed by the lexicon (Miikkulainen, 1993), and how the lexicon is acquired during development (Li, Zhao, & MacWhinney, 2007).

Following the Kroll & Stewart model, and using SOMs as its building blocks, the BiLex model consists of three separate maps: one for word meanings, and one each for phonetic symbols in L1 and L2, as illustrated in figure 1. All maps are linked by associative connections of varying strength, which allow network activation to flow between them.

Training Corpus During model training, the semantic and phonetic maps need to organize according to similarity, i.e. on the semantic map, words with similar meaning will tend to be close, while on phonetic maps, words that sound similar will tend to be close. For this organization to occur, semantic and phonetic symbols need to be encoded as vectors that

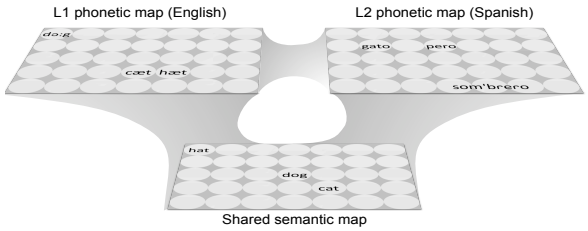

Figure 1: The BiLex model consists of three SOMs, one each for semantics, L1, and L2, that are linked by associations that enable the model to translate between semantic and phonetic symbols, simulating lexical access in bilingual humans.

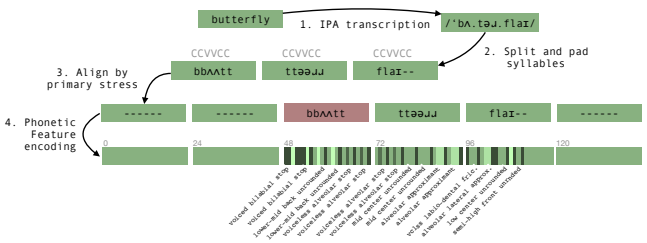

Figure 2: A word is encoded as a phonetic vector representation, creating the basis for phonetic map organization in the BiLex model.

reflect this similarity.

Feature-based semantic and phonetic vector representations were developed for a training corpus of 638 concrete nouns in English and their direct translations to Spanish. Semantic representations were derived from feature data developed by Sandberg, Gray, and Kiran (2018). For each word, 10-20 relevant attributes (e.g. "can fly") were used that were normed on healthy adults using Amazon MTurk. Overall, data from more than 320,000 interactions of the type "does word X have feature Y?" were used to produce semantic vectors of 400 features.

Phonetic representations were based on phonetic transcriptions of English and Spanish words, which were split into spoken syllables, and padded such that the primary stress lined up for all words. The individual phonemes comprising each syllable were represented numerically as a set of phonetic features like height and front-ness for vowels, and place, manner, etc. for consonants (Ladefoged, 2001). Figure 2 illustrates the encoding process. The final phonetic representations consisted of 144 real-valued features for English, and 192 for Spanish.

Model Training

Using the semantic and phonetic symbols as input data, the organization of the three maps and the associations between them are learned simultaneously. Symbols are presented to two of the maps at the same time; the two exposed maps adapt, and over time become more likely to represent each symbol in the corpus accurately. At the same time, associative connections between corresponding semantic and phonetic symbols grow stronger.

Varying relative exposure to English and Spanish can be

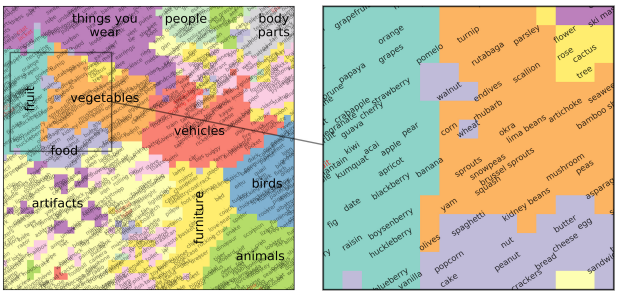

Figure 3: A well-trained semantic map, with winner units for each word labeled (LEFT). Map units are colored according to semantic categories, showing good global organization. RIGHT: Detail demonstrating that semantic similarity is reflected locally as well (e.g. walnut, nut, peanut are neighbors). Carefully designed training parameters are essential in creating this kind of highly organized map.

simulated by presenting English and Spanish phonetic symbols at proportional frequencies during model training, enabling the model to capture the effects of an individual's language learning history.

SOM Training Each SOM consists of a two-dimensional grid of neurons; each neuron is associated with a weight vector that encodes a semantic or phonetic symbol. The maps are trained using the standard SOM training algorithm (see e.g. Kohonon, 2001), which causes the weight vectors to become representations of the symbol vectors. At the same time, neighboring weight vectors become similar, and the map learns to represent the space of symbols as a two-dimensional layout where units that are close to each other on the map are similar either semantically (in the semantic map) or phonetically (in the phonetic maps).

SOM training is mainly governed by two parameters: the learning rate α determines the intensity of training, and the *neighborhood size* σ determines whether a larger or smaller part of the map changes in response to a training input.

The effectiveness of SOM training depends critically on how σ and α change over time. To develop the map's global structure first, the size of the neighborhood usually starts relatively large (on the order of the size of the map), and is gradually reduced, which causes the map to learn the similarity relations between input patterns at a more and more fine-grained level. Similarly, the learning rate is usually reduced over time, which allows the map to fine-tune its weight vectors in later stages of training.

Figure 3 shows an example of a well-trained semantic map, with colors encoding rough semantic categories. The categories tend form contiguous areas that border on similar categories. The detail on the right illustrates that locally, concepts tend to be arranged according to semantic similarity as well, e.g. "walnut", "nut", and "peanut" form a tight cluster.

A central working assumption underlying the BiLex model is that, similar to the training schedules necessary to achieve well-organized SOMs, language acquisition during human development requires an equivalent progression of factors governing learning. In other words, the cortical structures that underlie the human lexicon start out highly flexible and adaptive, but later in life adapt only to a smaller degree, both in terms of learning intensity and overall flexibility. In this way, SOM-based models can provide a mechanistic explanation for the age-related limitations on second language learning that occur in humans.

Training Associative Connections In addition to map training, associative connections between the maps are adapted simultaneously based on Hebbian learning, i.e. by strengthening connections that link active neurons:

$$a'_{ij} = a_{ij} + \alpha\theta_i\eta_i\theta_j\eta_j,$$

where a_{ij} is the weight of the associative connection from unit i in one map to unit j in the other map, η_i is the activation of unit i, and θ_i is a function defining the current map neighborhood.

In order to prevent the associative strengths from increasing indefinitely, the the overall sum of outgoing associative connections is normalized such that for each neuron, the L2 norm of outgoing connections to each target map is 1.

Additionally, since lexical access can decline in humans with age or lack of exposure to a language (Kavé, Knafo, & Gilboa, 2010), small amounts of random noise (with a given variance γ) are added to the associative connections during training.

Simulating Naming Tests Once a BiLex model is trained, the task of naming an individual word in either language can be simulated by first presenting its semantic representation to the semantic map. The resulting map activation is then propagated to the phonetic map via the associative connections; the weight vector of the most highly activated phonetic unit is then compared to all phonetic representations in the corpus, and the word with the minimal distance is produced as output. If the output word matches the original input, the word is counted as correctly named. The simulated naming performance for a set of words is the percentage of words that are correctly named in this way.

Evolutionary Parameter Fitting

In BiLex, age and relative language exposure over time are based on individual human data: the age of an individual determines the number of epochs used for model training, one training epoch per simulated year. The relative exposure to each language determines the proportion of English vs. Spanish words randomly selected for training during each epoch. However, appropriate settings for all remaining parameters governing the training process are initially unknown, including how learning rates and neighborhood sizes for the SOMs change over time. Finding parameter settings that enable BiLex to match an individual's naming performance given past language exposure is a complex problem, involving precise tuning of a large set of interdependent parameters. The

remainder of this section describes an Evolutionary Algorithm (EA; e.g. Bäck et al., 1997) designed to solve this problem.

EAs are a class of population-based optimization algorithms that use mechanisms inspired by biological evolution, like reproduction and mutation, to solve optimization problems. EAs maintain a population of candidate solutions, using a *fitness function* to determine the quality of each candidate. Highly fit candidates are more likely to be selected to reproduce, and recombine with other highly fit candidates. In this way, evolution tends to produce better candidate solutions over time.

Representation of Candidate Parameter Sets For the present problem of finding the best possible parameter settings for BiLex, each candidate solution was a set of BiLex training parameters, excluding age and relative language exposure, but including α and σ at different simulated ages, the scale γ of the random noise added to simulate aging and attrition effects, and the size N for each SOM.

To avoid overfitting, both α and σ were assumed to be the same for all three maps. Specific values for α and σ were evolved at a number of simulated ages (1,4,7,10,13,19,25, and 50 years), and interpolated linearly for intermediate values. Additionally, both α and σ were constrained to non-increasing values, i.e. at each time, the minimum of all values so far was used for training.

Training the associative connections also requires a learning rate α' at each time during training. To limit the number of parameters, a single factor k was added, such that at each time, $\alpha' = k \times \alpha$. In this way, the scale of α' was independent of that for SOMs, but changed in the same way over time.

To account for the fact that monolinguals tend to score above zero on naming tests in the other language, a minimum exposure parameter ε was added such that exposure for each language was clipped to values between ε and $1-\varepsilon$.

Initially, the number of words trained per simulated year were also included in the set of evolved parameters, which turned out to be unnecessary. In the reported experiments, the number of trained words per simulated year was set to a fixed value of 1.5 x the size of the training corpus.

Overall, each candidate parameter set was encoded using 20 numeric values; the initial population of 100 candidates was generated using random values within reasonable intervals, which were chosen empirically for each parameter. E.g. neighborhood sizes were constrained to an interval between 0 and 10, and initial learning rates ranged from 0 to 0.4.

Evaluation and Age-Layering In order to evaluate how well a particular candidate was able to match the naming performance of a given human participant, a BiLex model was trained, and the naming tests administered to the human participant were simulated using the trained model. The goodness-of-fit for a given candidate on a human individual i (GOF_i) was then calculated as the sum of squared residuals for the naming scores in both languages.

Based on this GOF measure, the straightforward way of fully evaluating the fitness of a candidate would be to evaluate it on all training samples, and compute the fitness as the mean GOF measure, requiring training and evaluating a complete model for each i, and making the evaluation function extremely expensive.

As a possible solution, age-layered EAs (Shahrzad et al., 2016) attempt to limit complete evaluations to only the most promising candidates. Candidates that score highly on an initial limited evaluation are further evaluated, while weak candidates are eliminated, saving computing resources.

Age Layering is particularly useful for noisy and expensive evaluations, and has been shown to speed up evolution significantly. To optimize BiLex parameters, a slight variation was used that accounts for the small, fixed set of human individuals on which each candidate can be tested: rather than ranking candidates by their overall fitness, a separate ranking for each human data set i was computed, and candidates were then discarded if their average ranking was below the 50th percentile within their age layer.

EA Design The remaining components of the Evolutionary Algorithm were fairly standard (see e.g. Bäck, 1997); Parents were chosen by standard roulette-wheel selection; offspring was created using uniform crossover, and mutated by adding normally distributed noise with uniform probablility (p=0.05) and standard deviation 0.025, scaled by the size of the initialization interval for each parameter.

In order to simplify distributed evaluations across remote compute nodes, and because age-layering makes the time required for evaluations unpredictable, a steady-state EA was used, i.e. rather that proceeding in distinct generations, population size was maintained between 50 and 70 candidates by adding new candidates continually as needed.

Finally, if none of the most recent 500 candidates was able to improve on the previous best solution, a mutation burst was performed, i.e. to maintain diversity, new candidates were added without recombination, but using a high mutation rate of 0.5. If no improvement was observed in the 1000 most recent candidates, the EA terminated, and the current best solutions were used as the final result. All parameters governing the EA were set empirically.

Experiments

Human Data The human data used to evaluate the parameter tuning methodology were collected from 33 healthy adult individuals, including 28 Spanish-English bilinguals and 5 monolinguals (2 Spanish, 3 English), who were included in order to provide the EA with appropriate edge cases w.r.t. language exposure and naming performance.

Relative exposure to English vs. Spanish over each individual's lifetime was estimated using a standard Language Use Questionnaire (LUQ19; Kastenbaum, 2018), which included questions about age, native and second languages, as well as a detailed self-reported linguistic profile that included relative exposure to both languages.

In order to measure lexical access (i.e. naming performance) in English and Spanish, all participants completed the Boston Naming Test (BNT; Kastenbaum, 2018), as well as another 60-item picture naming screener test used in clinical practice. To reduce the noise inherent in such tests, both tests were averaged to obtain one composite naming score for each language.

The provided data on language exposure and age made it possible to modulate relativve English vs. Spanish exposure over the course of the simulated lifetime for each individual human, creating an individual BiLex model whose naming performance could be measured and compared to the actual test scores.

Validating the Evolutionary Parameter-Fitting Method In order to evaluate the generalization performance of the proposed evolutionary method, a five-fold cross-validation run was conducted, using the human data described above as either training or test data. The initial set of 33 participants was divided randomly into five test sets, with each test set containing one monolingual. For each test set, the EA parameter optimization was performed using the remaining 26 or 27 healthy controls as training data. Generalization performance was measured as the goodness-of-fit on the respective test sets: For each individual in a test set, a model was trained using parameters that were evolved to fit the naming performance of the respective training set. Since each control subject was part of one test set, this was possible for all 33 controls.

Results

All five cross-validation runs produced highly fit candidate solutions; final best-fit parameter sets were found after evaluating 2749 (SD=1023) candidates on average, training and evaluating an average of 13549 individual BiLex models. Complete evaluation of all candidate parameter sets would have required over 7x as many trained models, suggesting that the age-layering approach was highly effective in reducing the number of required evaluations.

Most parameters in the best-fit candidate parameter sets tended to be similar, e.g. low but finite minimum exposure ε (0.04, SD=0.0137), and large initial neighborhood size (08.06, SD=1.17) that decreased dramatically (0.59, SD=0.049) by age 25.

Simulated composite naming scores were highly predictive of human data for both English ($R^2 = 0.77$, $p << 0.0001$) and Spanish ($R^2 = 0.63$, $p << 0.0001$). Figure 4 shows predicted vs. actual composite naming scores for both languages, using predicted naming scores from the top five parameter sets found by each of the five EA runs.

Figure 5 illustrates the way in which L2 age of acquisition (AoA) and exposure influence the structure of BiLex maps using concrete phonetic maps from four individual BiLex models; each map was trained using evolved training parameters and the language history of one of the bilingual study participants. The individual maps were chosen to represent

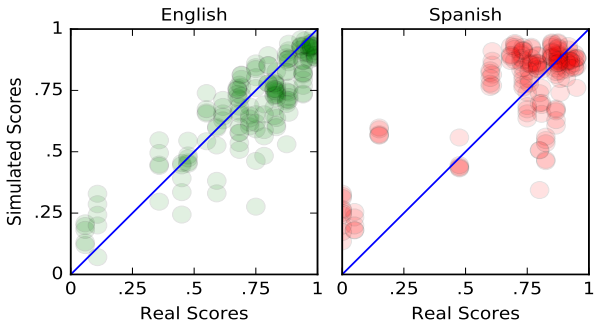

Figure 4: Simulated naming scores on test sets are highly predictive of human data for both English (left, $R^2 = 0.77$) and Spanish (right, $R^2 = 0.63$), indicating that models trained with evolved parameters are able to generalize to simulate bilingual access of previously unknown individuals.

extreme AoA/exposure combinations: Panel A shows early AoA and high exposure, from a model with high L2 naming score ($> 90\%$). Panel B demostrates that as long as the AoA is early, the L2 map organizes and performs well even for low exposure. Panel C shows a late-AoA/high exposure map; the global organization deteriorates to some degree, but performance is still acceptable (70%). Finally, panel D shows how the combination of late AoA and low exposure leads to a badly organized map that accounts for low performance ($<40\%$).

Discussion

The complexity of the BiLex model, the infinite possible combinations of individual language history, and the comparatively small amount of human data available in this case make BiLex an appropriate test case for the evolutionary parameter fitting method proposed in this paper. The reported results demonstrate clearly that using evolution, a complex model like BiLex can be configured to capture complex interactions between environment and behavior, in a model that itself plausibly models neural information processing.

In addition to capturing the link between language exposure and naming ability quantitatively, the same link was visible in the organization of phonetic L2 maps in the optimized model: either early L2 acquisition or high exposure lead to well-organized L2 phonetic maps and high naming performance, while low exposure and late acquisition led to deficient map organization and naming ability.

In this way, models based on known theories, and designed to account for quantitative data on a more abstract level, can still provide additional insight and generate unexpected explanations for mechanisms underlying a given phenomenon – in this case, about the way in which AoA and exposure modulate lexical acces through phonetic map organization.

Note that, while BiLex was used as a concrete example throughout, the method extends to similar models, and aims to make parameter fitting of complex neural network models to limited human data workable in general.

Finally, while evolution can help models such as BiLex explain normal human cognition and capture the ways in which

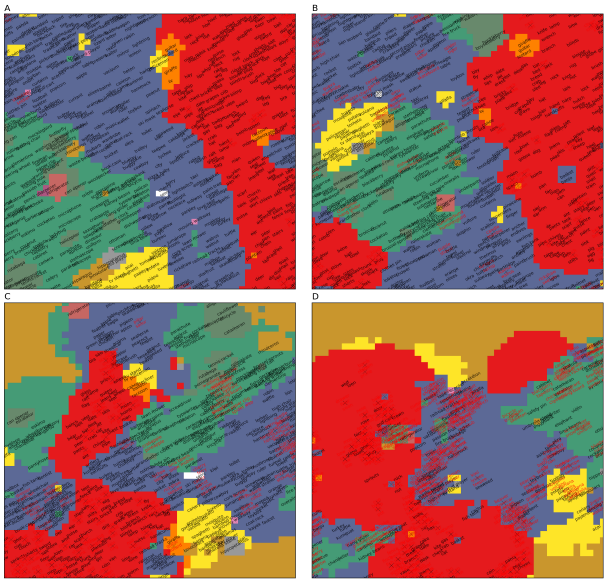

Figure 5: L2 phonetic maps of individual EA-optimized BiLex models. (A) Early age of acquisition (AoA) and high exposure leads to well-organized L2 phonetic maps. (B) Early AoA leads to well-organized maps despite low exposure. (C) Late L2 AoA impacts both global organization of the phonetic map even at high exposure. (D) Late AoA and low exposure lead to deficient global and local map organization. Taken together, the maps offer a mechanistic explanation for AoA/exposure effects seen in humans.

underlying brain mechanisms, environment, and cognitive function interact, the resulting models of normal cognition can also serve as a basis to investigate how these functions break down, and potentially inform the development of improved diagnostic methods and clinical interventions.

In current research, EA-optimized BiLex models are used in this way to create individual models of bilingual patients suffering from Aphasia; the resulting pre-morbid patient models are then used to simulate the onset of Aphasia, and to predict outcomes of alternative interventions. The approach is currently evaluated in an ongoing clinical trial, making it (to our knowledge) th first time a neural network model has been used in this way – the systematic, mechanical way of optimizing the model that was introduced in this paper makes novel modeling application such as these possible.

Conclusions

This paper proposed an evolutionary approach designed to make fitting complex NN-based models of higher cognition to limited data workable in practice. An Evolutionary Algorithm was introduced and evaluated by optimizing training parameters for BiLex, a connectionist model of the bilingual lexicon. Using EA-optimized parameters, BiLex was able to capture the complex interactions between exposure to different languages and the resulting individual differences in bilingual lexical access, demonstrating how evolution can help build the next generation of computational models of cognition.

References

Abadi et al. (2015). *TensorFlow: Large-scale machine learning on heterogeneous systems.* (Software available from tensorflow.org)

Bäck, T., Fogel, D., & Michalewicz, Z. (1997). *Handbook of evolutionary computation*. Oxford Univ. Press.

Bhatia, T. K., & Ritchie, W. C. (Eds.). (2005). *The handbook of bilingualism*. Blackwell Publishing.

Caramazza, A., Hillis, A., Leek, E., & Miozzo, M. (1994). The organization of lexical knowledge in the brain: Evidence from category- and modality-specific deficits. In L. Hirschfeld & S. Gelman (Eds.), *Mapping the mind.* Cambridge University Press.

Kastenbaum, J. G. e. a. (2018). The influence of proficiency and language combination on bilingual lexical access. *Bilingualism: Language and Cognition*, 1–31.

Kavé, G., Knafo, A., & Gilboa, A. (2010). The rise and fall of word retrieval across the lifespan. *Psychology and Aging*, *25*(3).

Kohonen, T. (2001). *Self-organizing maps* (3rd, extended ed.). Berlin: Springer.

Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language*, *33*.

Ladefoged, P. (2001). *Vowels and consonants: An introduction to the sounds of languages*. Oxford: Blackwells.

Li, P., Zhao, X., & MacWhinney, B. (2007). Dynamic self-organization and early lexical development in children. *Cognitive Science*, *31*.

Maass, W., Natschlager, T., & Markram, H. (2002). Real-Time Computing Without Stable States: A New Framework for Neural Computation Based on Perturbations. *Neural Computation*, *14*(11).

Miikkulainen, R. (1993). *Subsymbolic natural language processing: An integrated model of scripts, lexicon, and memory*. MIT Press.

Sandberg, C. W., Gray, T., & Kiran, S. (2018). Development of a free online interactive naming therapy for bilingual aphasia. In *American speech language hearing association convention.*

Schmidhuber, J. (2014). Deep learning in neural networks: An overview. *CoRR*, *abs/1404.7828*.

Shahrzad, H., Hodjat, B., & Miikkulainen, R. (2016). Estimating the advantage of age-layering in evolutionary algorithms. In *Proceedings of the genetic and evolutionary computation conference (gecco-2016, denver, co).*

Spitzer, M., Kischka, U., Gückel, F., Bellemann, M. E., Kammer, T., Seyyedi, S., … Brix, G. (1998). Functional magnetic resonance imaging of category-specific cortical activation: Evidence for semantic maps. *Cognitive Brain Research*, *6*.

Theano Development Team. (2016). Theano: A Python framework for fast computation of mathematical expressions. *arXiv e-prints*, *abs/1605.02688*.