Transfer Learning Across Heterogeneous Tasks Using Behavioural Genetic Principles

Maitrei Kohli
Dept. of Computer science & Information Systems
Birkbeck University of London
London, United Kingdom
maitrei@dcs.bbk.ac.uk

George D. Magoulas
Dept. of Computer science & Information Systems
Birkbeck University of London
London, United Kingdom
gmagoulas@dcs.bbk.ac.uk

Michael S.C. Thomas
Dept. of Psychological Sciences
Birkbeck University of London
London, United Kingdom
m.thomas@psychology.bbk.ac.uk

Abstract **We explore the use of Artificial Neural Networks (ANNs) as computational models capable of sharing, retaining and reusing knowledge when they are combined via Behavioural Genetic principles. In behavioural genetics, the performance and the variability in performance (in case of population studies) stems from structure (intrinsic factors or genes) and environment (training dataset). We simulate the effects of genetic influences via variations in the neuro-computational parameters of the ANNs, and the effects of environmental influences via a filter applied to the training set. Our approach uses the twin method to disentangle genetic and environmental influences on performance, capturing transfer effects via changes to the *heritability* measure. Our model captures the wide range of variability exhibited by population members as they are trained on five different tasks. Preliminary experiments produced encouraging results as to the utility of this method. Results provide a foundation for future work in using a computational framework to capture population-level variability, optimising performance on multiple tasks, and establishing a relationship between selective pressure on cognitive skills and the change in the heritability of these skills across generations.**

Keywords—transfer learning, behavioural genetics, artificial neural networks, heritability, genetic algorithms

I. Introduction

Transfer learning is a research field in machine learning which aims to store and reprocess the knowledge gained while learning one task to learn different but related tasks [1]. This concept draws inspiration from the psychological notion of *transfer*, which explores how enhancement in one mental function could influence another related one. Machine learning methods that enable any kind of communication between different tasks are performing transfer. The task from which the knowledge is extracted is called the *source* task and the novel task to which it is applied is the *target* task. Literature in the field organises methods for performing transfer into two distinct types – functional and representational [2]. In functional transfer, learning in the source and target happens simultaneously and it exploits implicit pressures from additional training patterns, via shared or common internal representations. In representational transfer, source and target learning occurs separately in time and an explicit representation is transferred from the source to the target. It cares most about learning the target task only.

Most methods of transfer learning implicitly assume that the source and target tasks are somehow related to each other - when, for example, the source task concerns training on female-only speech whilst the target task is to recognise speech from males only. In addition, most existing transfer learning algorithms assume that the feature spaces between the source and target domains are the same. However, in practice, it is useful to transfer knowledge across domains or tasks that have different feature spaces - the so-called *heterogeneous transfer learning* [1].

If the assumption about relatedness does not hold true, transferring knowledge might result in negative transfer. Negative transfer refers to the impairment of current learning and performance due to the application of non-adaptive or unsuitable information. Estimating task relatedness and finding ways to avoid negative transfer pose a challenge which is attracting the attention of researchers in the field. Many methods have been reported to assess task relatedness. One commonly used framework to consider transfer is using artificial neural networks (ANNs). For ANNs, the most commonly used method to assess relatedness is the distance between weight space representations – the smaller the distance, the stronger the task relatedness. Another open question is which part of knowledge to share or transfer across tasks. For ANNs, most commonly used techniques involve the use of shared weights, common hidden layer, global learning rate and common training set [1][3][2][4]. Another important consideration is developing algorithms/methods capable of transferring this knowledge. Examples include Bayesian models, Hierarchical transfer methods, Relational transfer methods as described in [1][5] and references therein. Various attempts have been made to overcome the aforementioned challenges. Many are successful, albeit within their limited scope or case-specific applications. However, there is an increasing need for transfer learning techniques used for broader and more challenging applications. This in turn requires having more generalised methods that can be applied on any given set of tasks.

In this work, we address the following key challenges: to perform heterogeneous transfer, avoid negative transfer, and propose a mechanism for determining task relatedness which extrapolates well to different domains and embodies the effects of both structure/intrinsic parameters and training datasets within which the learning system is placed. To this end, we propose a novel transfer approach to learn multiple heterogeneous tasks using concepts drawn from Behavioural Genetics [6][7]. Research in this multidisciplinary field shows that performance is highly dependent on both the genes and the environment. We draw an analogy between genes and intrinsic parameters, and the training dataset and the environment. Within Behavioural Genetics, it is well known that the quality of environment can modulate the influence of genetic variation. Following the analogy, one can similarly observe that training datasets affect the influence of intrinsic parameters. Thus, for ANNs, a certain number of hidden units may be highly beneficial for a specific condition of the dataset (say, for the number of training examples available) but if these conditions were to change drastically, the same number of hidden units may no longer be optimal. Thus, the system's performance will alter.

The proposed approach combines concepts of Behavioural Genetics with the idea of a parametrically diverse populations of learning systems, used in the context of a hybrid genetic algorithm, where genes (representing intrinsic factors) and environment (expressed via training datasets) interact throughout development to shape differences in individual classifier behaviours (performance). The method spans transfer learning systems and multi-task learning systems, incorporating "good/useful" features of both, and then combines them with principles of Behavioural Genetics. A comparative analysis of the differences between the proposed transfer approach and others closely related in the literature is presented in Table 1.

The rest of the paper is organized as follows: in Section II we present the proposed behavioural genetics based approach and highlight some of the key aspects; in Section III we describe the methodology and its implementation. Section IV describes the experiments and presents the results. In Section V we discuss the results; and, finally, the paper ends with conclusions and future work in Section VI.

II. Behavioural Genetics-based Model for Learning Transfer

In this work, the effects of genetic influences are simulated via variations in the neuro-computational parameters of the ANNs. These parameters relate to how a network (an individual of the population) is built, its processing dynamics, how it is maintained, how it adapts and how it generates behavioural outputs. The effects of environmental influences are simulated via a filter applied to the training set. The filter creates a unique subsample of the training set for each simulated individual in the population, inspired by the notion of socio-economic-status (SES) in Behavioural Genetics [8][9]. Our approach uses a population of twins (ANNs with some degree of similarity in their neuro-computational parameters) to disentangle genetic and environmental influences on performance. The model captures the wide range of variability exhibited by population members as they are trained on and across different tasks. This approach is inspired by cognitive development, where twins are more closely matched for age, family and other social influences. This is because twins are either genetically identical (genetic relatedness of 1.0 for *mono-zygotic,* MZ, or identical twins) or as similar as siblings (genetic relatedness of 0.5 for *di-zygotic,* DZ, or fraternal twins) and, to an approximation, share the same environment (applicable for both MZ and DZ twins based on the *Equal Environment assumption*) [10]. The difference in the similarity in performance between MZ or DZ twin pairs, along with assumptions about their similarity of environment, allows inferences to be drawn about the influence of genetic relatedness on behaviour [11].

From a computational point of view, in particular we can exploit the notion of *heritability* within Behavioural Genetics to assess task relatedness. Heritability is a statistic that describes the effect size of genetic influence and refers to the proportion of observed or phenotypic variance that can be explained by genetic variance. In simpler terms, it is the amount of population variability explained by genetic similarity [11]. In computational terms, heritability can be interpreted as the amount of performance variation accounted for by structural similarity. Twin studies provide an exact computation of heritability.

	Learning Goal	Type of Transfer	Degree of task relatedness	Means of assessing task relatedness	Special features
Multi task Learning	Improving performance in all tasks	Functional	Highly interrelated	Case-specific relatedness measures	Involves use of shared internal representations such as weights, common data sets for all tasks
Transfer Learning	Improving target task performance	Representational	Related but may be from different domains	Application/case/domain specific measures only but cannot be applied per se on a generalised basis	Highly application/case sensitive
Behavioural Genetics based	Improving performance in all tasks	Hybrid: works sequentially like representational and uses common internal representations of intrinsic parameters, like the functional	Can be unrelated or heterogeneous	Heritability and change in heritability, can be used in any scenario	It is based on principles of Behavioural genetics; incorporates shared intrinsic parameters and effects of environment on performance (epigenetics)

Table 1: Comparison between the proposed approach and other related approaches

Additionally, twin studies provide a valuable tool for exploring environmental influences, especially family or shared environment, against a background of heritability. Since twins are genetically similar, if heritability affects behaviour then MZ twins will be more similar than the DZ twins. In other words, heritability (with twin studies) provides an estimate of the magnitude of genetic influence on behaviour [10].

Heritability is an integral part of the current work for following reasons. As a population is bred and optimised across generations on a particular task, the range of variation of its *suitable or relevant* computational (or intrinsic) parameters reduces, i.e. optimisation leads to reduction in heritability. If the range of environmental variation is kept the same, the variation in performance will be more due to environmental variation, since the optimised population will now be more genetically homogeneous. Now, consider if the same population were trained on another non-related or heterogeneous task, and this also experienced a reduction in heritability. This would be an indication of the presence of some kind of relatedness among the tasks. The *direction of the change in heritability indicates task relatedness*. If the change in heritability for different tasks is in the same direction (e.g. all values decrease or all increase proportionally), this implies that the same set of intrinsic parameters are appropriate for learning the tasks. This in turn can help in identifying a set of *domain-relevant* parameters, which, like generalist genes [12], are useful for learning various heterogeneous tasks. Thus, change in heritability has the potential to act as a mechanism for identifying task relatedness, which extrapolates to different task domains, and consequently avoids negative transfer.

III. Implementation of the Approach

The various steps involved in implementing our methodology are outlined below:

- The first step is to identify *n* number of heterogeneous tasks. In our experiments below, $n = 5$. The five tasks were: English past tense, autoassociation, consistent categorisation, categorisation with exceptions, and arbitrary association. The tasks were chosen to vary with respect to their characteristics such as degree of similarity between the input and the output patterns, the presence of structure or regularity in mappings and the overall complexity. Thus, all five tasks pose different requirements to the ANNs.
- For each of the five tasks there are two datasets: one is used for training and the other one is used for calculating the generalisation accuracy. For this instantiation of the framework, all training and generalisation/test datasets have 57 bit inputs and 62 bit outputs representing different types of features from the five tasks. Table 2 provides the description of all tasks and their training datasets. The first task, English past tense formation, was included because the framework was first utilised as a psychological model of individual difference in language development [13]. Features of this task represent verb phonemes whilst, for example, features of the categorization task represent attributes of patterns belonging to 10 categories. The same is applicable to categorisation with exceptions task also but with the exception that patterns belonging to category 9 whose Euclidean distance to the prototype pattern is less than 2 are assigned to category 7 instead. In case of auto association task, the patterns were 57 bit random vectors which were same for inputs and targets. In arbitrary mappings, the inputs and targets consisted of completely different 57 bit random patterns.
- Choose any one task as the source task, and the remaining *n*-1 tasks become the target tasks. The aim is to successfully learn all tasks.
- Encode the neuro-computational parameters of artificial neural networks in a genome. The parameters used in this work which are encoded within a fixed range include number of hidden units, learning rate, slope of logistic activation function, weight decay and nearest neighbour threshold. This stipulates the range of variation for all neuro-computational parameters. In this paper the range is chosen with the intent to help learn the source task better. Thus, each member of the population will have a different set of values, but within the same chosen range for the encoded parameters ensuring genetic diversity.
- A population of 50 pairs of MZ twins and 50 pairs of DZ twins is created by simulating the biological processes of meiosis and fertilisation. In this work, as we progress with the generations, only offspring are included in the new generation populations.
- In the current implementation, the environment is represented by training sets. Environmental variability is implemented as a filter applied to the training tasks, inspired by research on how SES affects cognitive development. A body of research suggests that individuals in lower SES families experience substantially less quality and quantity of information [8]. The filter creates a unique subsample of the training set for each simulated individual, based on a parameter determining the quality of the environment. An individual's environmental quality is modeled by a number selected at random from the range 0.6-1.0. This gives a probability that any given pattern in the full training set would be included in that individual's training set. This filter is applied at each generation to create unique training subsets for all members of the population in that generation. The range 0.6-1.0 defines the range of variation of environmental quality, and ensures that all individuals are exposed to more than half of the training dataset. Due to the *equal environment assumption*, twin pairs have the same training subset.
- The population of twins, *twinpop1* and *twinpop2* are then trained on the source task and independently on each of the target task.

- Performance assessment and Heritability has two parts: calculating classification accuracies, including generalisation performance and measuring heritability and population variability. This step is done at the end of training for each generation. Performance is assessed using recognition accuracy based on *Hamming distance.* Heritability is measured at the end of each generation using *Falconer's equations* [10]
- Measuring heritability involves calculating MZ and DZ correlations. This is done by using the Pearson correlation formula

$$\rho_{x,y} = \mathrm{cov}(x,y)/\sigma_x\, \sigma_y$$

where, x and y are performance vectors of pairs of MZ, or DZ, twins sampled from the twinpop1 and twinpop2 respectively, and σ_x denotes variance in x; σ_y is the variance in y.

Tasks	Input Bits	Output Bits	Description of data set
Modelling performance of 6 year old children on English Past Tense	57	62	a) The training set consists of 508 English past tense verbs; each verb split in 3 phonemes (19 bits each) b) Type frequency of verbs: 410 – regular, 20 – identical, 68 – vowel change, 02 – arbitrary c) 8 arbitrary (non English) verbs for ensuring finer graduations of performance. d) Token frequency implemented by multiplying the corresponding bit with change in weight due to difference between actual and the target output. e) Separate test set consists of 500 novel verbs
Auto association	57	62	a) Training set consists of 500 patterns and Target patterns same as input patterns b) ANNs produce 62-bit output vectors (62 output nodes) but the last 5 bits get zero values for all mappings. c) Separate test set consists of 500 novel patterns.
Consistent Categorisation	57	62	a) Training set consists of 500 patterns belonging to 10 categories b) Each pattern is assigned a category based on its similarity to the prototype pattern of each category. c) Each pattern is created by altering each bit of corresponding prototype pattern with a probability of 0.05 d) Test set consisting of 500 novel patterns using the same procedure used for test set.
Categorisation with exceptions	57	62	a) Training set consists of 500 patterns where same input patterns as in previous categorisation data set are used. b) Slight modification in the mappings. Includes a sub cluster of exceptions. c) This sub cluster consists of all input patterns of category 9 whose Euclidean distance from prototype element of the category is less than 2. d) These patterns are assigned category 7, instead of 9. e) Test set consisting of 500 novel patterns using the same procedure used for test set.
Arbitrary Association	57	62	a) Training set consists of 500 patterns ; targets are not same as the inputs b) No generalisation set since random inputs have random outputs.

Table 2: Heterogeneous tasks and dataset description.

- The formula for computing heritability is

$$h^2 = 2\,(r_{MZ} - r_{DZ})$$

where h^2 represents additive genetic effect (or narrow sense heritability); r_{MZ} is the MZ correlation and r_{DZ} is the DZ correlation.

- The proportion of variance due to shared environmental influences (filtered training sets) is calculated

$$c^2 = r_{MZ} - h^2$$

where c^2 represents proportion of variance due to shared environmental influences; r_{MZ} signifies MZ correlation and h^2 is the heritability.

- Lastly, the proportion of variance due to non-shared environmental influences (stochastic factors unique to each individual) is calculated as

$$e^2 = 1 - r_{MZ}$$

where e^2 embodies proportion of variance due to shared environmental influences and r_{MZ} refers to MZ correlation.

- Based on the performance of the population of networks on source task, members are *selected* from twinpop1 only for breeding the next generation. The selection criterion used is the standard roulette wheel selection which is applied at the end of training (1000 epochs).
- This process is iterated until ANN parameters do not change any more significantly or performance starts converging, i.e. learning error in tasks reaches small value.

IV. EXPERIMENTS AND RESULTS

The results we report follow four generations that were increasingly optimised on the past-tense task. We trace the change in performance across generations on this task, and the change in heritability; but also, crucially, we report the same measures when each succeeding past-tense-optimised generation is instead trained on one of the other four tasks. We report the results from training 100 pairs of feed forward neural networks using batch version of RProp algorithm. The stopping condition was an error goal of 10^{-5} within 1000 epochs. The reported results are for generations 1 to 4 (G1, G2, G3, G4) based on a population of 100 twin pairs, characterised as: Twinpop1, which is the population containing 1st twin out of each twin pairs (100 networks) and Twinpop2, which is the population containing the remaining 2nd twin out of a twin pair (again 100 networks). There were 508 patterns in English past tense training set and 500 patterns in the training set for the remaining tasks. In addition, there was a separate test dataset consisting of 500 novel patterns for each task in order to assess the generalisation performance of the networks. *The networks were trained on the filtered training sets but the performance was always assessed on the full training set and then tested on the previously unseen generalisation set.* Table 3 shows mean classification accuracy or the performance on full training set and Table 4 depicts the mean generalisation accuracy achieved by the population on each task across four generations. Heritability, effects of shared and non-shared environmental influences on each of the five tasks across four generations are shown in Table 5. Figure 2 contains the graphical representation of emerging heritability trends for all five tasks across the four generations.

In order to compare our approach with classical transfer approaches and as a baseline to assess transfer effects against, we used the method suggested in [14] to train randomly initialized networks and networks initialized by the literal transfer method. The comparison experiments begin by initialising 10 networks, each with random initial weights. These are the source networks and are trained on the source task: English past tense for 1000 epochs using Rprop algorithm and different randomly generated values of hidden units [within the range: 50-500] and learning rate [within the range: 0.1-0.9]. Performance is assessed at the end of the training and the trained network's weights are saved. The four remaining tasks become the target tasks. From each target task training dataset, 10 different 90%/10% holdout conditions were generated. Networks were trained on the 90% partition and the generalisation ability was tested on previously unseen 10% partition of target task dataset. For each holdout, 10 different randomly initialised networks were used for target training. We first check the case of 'random networks'. There are 10 randomly initialised networks per target task data holdout, i.e. they start training with random initial weights. These are trained using Rprop algorithm for 1000 epochs. Performance is measured at the end of the training. The other technique is the 'literal transfer' wherein the literal weights from each one of the source networks were transferred. The other conditions were the same as those in the former case. Thus, there were 100 different training runs per holdout and per task. The essence of doing this comparison is that – if the results obtained by our proposed approach are influenced by the inherited parameters and training datasets rather than due to chance, one would not expect to see similar changes if these networks were run under these comparison conditions.

V. Discussion of Results

Our experiments demonstrate several key points. The range of neuro-computational parameters was calibrated with respect to the *English past tense* acquisition task, thus making it the *main* or *source* task. Tables 3 and 4 show that overall performance on the full training set and test/generalisation set is good for all five tasks. As expected, performance on English past tense consistently improved over generations.

The performance drops drastically for association tasks in generation 2. The rationale behind this poor performance on the association tasks is that networks in generation 2 have genes (or intrinsic parameters) more suited to English past tense (since classification accuracy on that task was the selection metric). Whilst selecting members (from generation1) for breeding, the networks with genes more suited for English past tense task got selected and therefore the offspring or generation 2 networks automatically have better propensity towards these types of tasks. In addition, networks consider past tense, categorisation and categorisation with exceptions as belonging to similar domain (since they all classify patterns into one of the many classes). Hence, networks optimised on past tense perform well on the other *related* tasks too, because all three tasks pose same intrinsic requirements. On the other hand, the networks fare poorly on association tasks since the parameters being optimised (for past tense) are not suited for these tasks i.e. the learning mechanism considers that these tasks belong to non-related domain (as compared to the aforementioned three tasks).

Table 5 shows the twin correlations, heritability values and the effects of shared and non-shared environmental influences on the variability in performance. As expected MZ correlations are very high and DZ correlations are low. The heritability values in some generations are very high - more than 1. Since heritability is defined as the proportion of population variability explained by genetic relatedness, its value cannot logically exceed 1. Technically, values greater than 1 occur when the genetic effects are so strong that they violate the additivity assumed by the Falconer model. However, for our purposes, it is still meaningful to use the metric to compare across conditions, since the computed heritability remains proportional to the greater performance similarity of MZ twins over DZ twins, and therefore the net effect of the genetically determined parameter sets on population variability.

The heritability analysis from Figure 1 also illustrates the trend wherein: variations in heritability are in the same direction for the English past tense, the categorisation and the categorisation with exceptions tasks, which seem to form one group, and the auto and arbitrary associations as the other group. The direction of change in heritability values provides useful insight into assessing relatedness amongst tasks. If the direction of change is the same for different tasks, irrespective of whether it is positive or negative, it indicates that these tasks are related.

A key advantage of using heritability as a metric of task relatedness is that it summarises the net effect of all computational parameters varying within the learning system. As the heritability statistic measures variation in the performance values, the method is robust to increases in the number of parameters that vary in the learning systems, and which underlie any transfer effect.

	Min no of patterns (%)				Max no of patterns (%)				Mean (%)				STD(%)			
	G1	G2	G3	G4	G1	G2	G3	G4	G1	G2	G3	G4	G1	G2	G3	G4
English Past Tense [*Source*]																
Twinpop1	1.1	35.2	40.9	9.4	88.5	88.5	89.5	88.1	77.1	77.4	78.5	81.2	14.1	11.3	10.0	14.3
Twinpop2	3.7	37.4	44.2	23.2	89.1	88.3	89.5	87.9	77.9	77.6	78.6	82.1	12.8	11.1	10.0	10.9
Auto association																
Twinpop1	4.8	2.6	11.6	16.8	100	100	100	100	95.0	72.5	74.5	96.7	18.7	28.3	26.3	13.3
Twinpop2	6.2	4.0	36.4	6.2	100	100	100	100	96.8	69.79	74.4	96.5	15.0	28.2	25.3	14.7
Arbitrary association																
Twinpop1	4.6	5.2	2	1.6	59.4	59.8	74	74	45.6	45.5	48.8	37.2	11.9	12.2	19.9	23.0
Twinpop2	2.4	7.6	2.4	4.2	59.4	59.6	74.2	73.8	45.9	45.4	46.9	39.2	11.6	11.5	21.9	21.4
Categorisation																
Twinpop1	11.6	72.6	79.4	11.6	100	100	100	100	95.5	97.8	97.5	95.0	19.1	4.7	4.6	19.8
Twinpop2	11.6	76.8	33	11.6	100	100	100	100	97.2	97.4	96.8	95.5	14.6	4.9	7.9	19.3
Categorisation with Exp.																
Twinpop1	84.8	61.4	78.4	11.6	99.4	99	99	98.6	95.9	95.6	95.8	92.8	3.5	5.6	4.3	20.3
Twinpop2	86.8	68.8	39.8	11.6	99.6	99.2	99	99.4	95.7	95.3	95.6	93.7	3.4	5.5	6.7	18.9

Table 3: Classification performance of twinpop1 and twinpop2 on full training sets for generations G1-G4.

	Min (%)				Max (%)				Mean (%)				STD(%)			
	G1	G2	G3	G4	G1	G2	G3	G4	G1	G2	G3	G4	G1	G2	G3	G4
English Past Tense[*Main Task*]																
Twinpop1	3.0	28.9	27.7	12.2	91.3	90.9	89.3	94.8	67.72	70.97	70.96	82.78	18.75	15.58	14.52	15.65
Twinpop2	8.6	23.2	27.9	6.6	92.1	94.2	90.7	93.8	69.65	70.36	70.66	83.11	16.17	15.65	15.11	13.75
Auto association																
Twinpop1	2.8	2.4	2.2	4.4	100	100	100	100	94.53	80.06	73.01	95.76	19.99	32.01	37.28	17.12
Twinpop2	2.4	3.0	2.0	4.8	100	100	100	100	97.13	78.77	73.58	96.75	14.45	32.62	36.73	14.44
Arbitrary association	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
Categorisation																
Twinpop1	9.4	65.4	65.2	9.4	100	100	100	100	95.43	96.15	95.39	94.85	19.70	7.71	8.36	20.55
Twinpop2	9.4	65.0	29.0	9.4	100	100	100	100	97.07	95.57	94.88	95.46	15.27	8.32	10.26	19.84
Categorisation with exceptions																
Twinpop1	66.0	58.8	64.8	9.4	100	100	100	100	90.58	94.80	94.18	94.22	9.00	9.04	7.87	21.36
Twinpop2	70.4	62.6	32.6	9.4	100	100	100	100	90.61	94.08	93.80	95.17	8.47	9.18	9.76	19.66

Table 4: Generalisation performance of twinpop1 and twinpop2 after each generation G1-G4 in each one of the task. Asterisk, *, indicates that no test set was used since random inputs produce random outputs.

	English P.T.	**Auto-association**	**Arbitrary mappings**	**Categorisation Task**	**Categorisation with Exceptions**	**Generations**
MZ correlation	0.96	0.91	0.63	0.99	0.86	G1
	0.96	0.99	0.99	0.92	0.91	G2
	0.95	0.99	0.99	0.9	0.90	G3
	0.96	0.99	0.99	1.00	0.99	G4
DZ correlation	-0.08	0.21	0.57	-0.029	-0.21	G1
	0.69	0.49	0.43	0.59	0.72	G2
	0.11	0.29	-0.002	0.24	0.28	G3
	0.61	0.36	0.39	0.61	0.56	G4
Heritability, $h^2 = 2 * (rMZ – rDZ)$	2.11	1.39	0.11	2.05	2.15	G1
	0.54	1.01	1.11	0.65	0.38	G2
	1.67	1.40	2.00	1.46	1.28	G3
	0.69	1.26	1.19	0.77	0.87	G4
Shared environmental effects $c^2 = rMZ - h^2$	-1.14	-0.48	0.52	-1.05	-1.29	G1
	0.41	-0.01	-0.12	0.26	0.52	G2
	-0.72	-0.40	-1.00	-0.48	-0.32	G3
	0.26	-0.26	-0.19	0.22	0.12	G4
Non-shared environmental effects $e^2 = 1 – rMZ$	0.03	0.08	0.36	0.0008	0.13	G1
	0.03	0.002	0.002	0.076	0.08	G2
	0.04	0.002	8.00e-04	0.022	0.09	G3
	0.03	0.004	0.005	0	0.0002	G4

Table 5: Values for Heritability, shared and non-shared environmental influences for each task across the four generations.

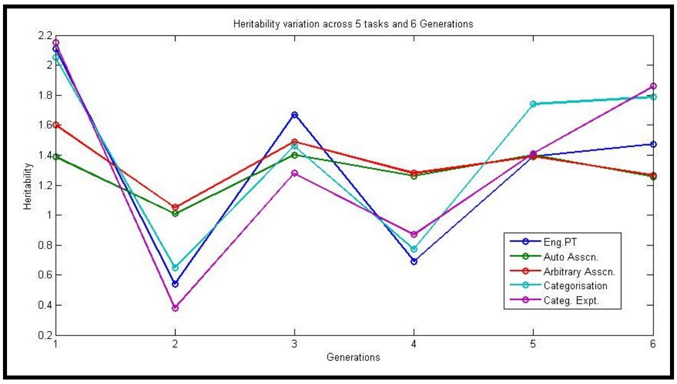

Figure 1: Heritability variation across generation.

Networks	Performance (%)
N1	61.02
N2	66.54
N3	75.79
N4	85.83
N5	81.89
N6	79.53
N7	63.98
N8	79.72
N9	64.57
N10	80.70

Table 6: Average performance per holdout in the source for random networks

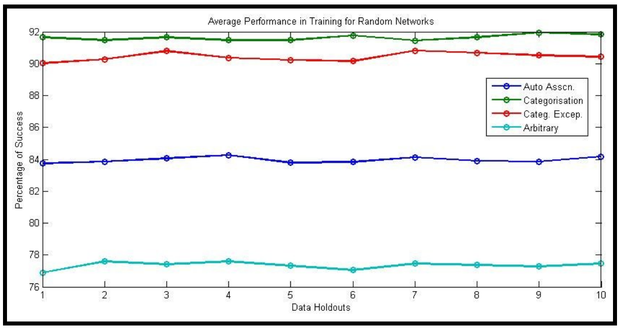

Figure 2: Average performance in Training for random networks

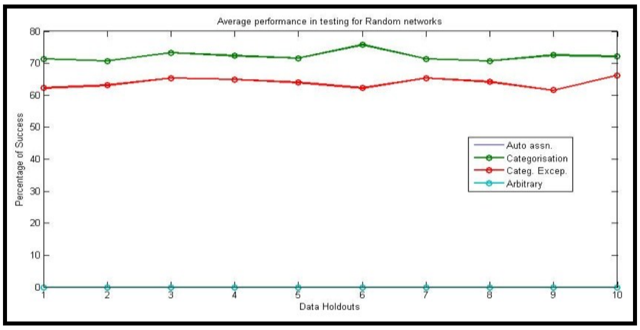

Figure 3: Average performance in testing for random networks

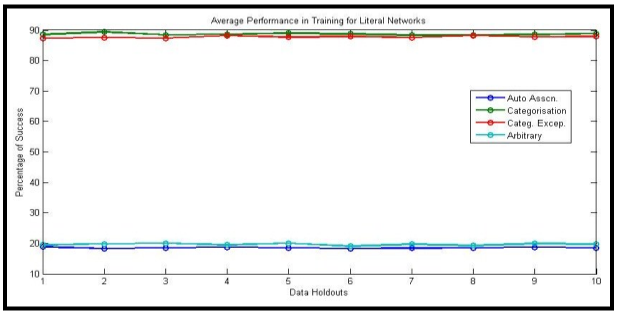

Figure 4: Average performance in training for literal networks

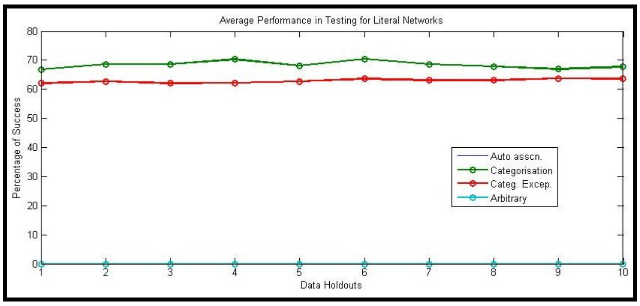

Figure 5: Average performance in testing for literal networks

Even though two distinct categories/clusters emerge, the performance improves for all 5 tasks. Shared knowledge (genome + filtered training set) thus helps all tasks, i.e. it acts in a domain relevant capacity.

Table 6 and Figures 2-5 depict the performance of the source and target tasks' on comparative approaches respectively. For the target task graphs, the average of all 10 networks per holdout was calculated and used. These graphs show that the average performance in training for randomly initialised networks is good on all tasks. However, when assessing the generalising capability of these networks, the performance on auto association and arbitrary mappings falls to zero, indicating that the networks are not capable of generalising these tasks at all. Even the generalisation performance of categorisation and categorisation with exceptions is not as good as the results achieved via our proposed approach.

When applying the literal transfer technique, the performance of auto association and arbitrary mappings falls compared to the randomly initialised networks; whereas the training performance on categorisation and categorisation with exceptions is very good. The generalisation capability is not very impressive in case of categorisation and categorisation with exceptions but it is non-existent in case of other two tasks. This further strengthens our conclusion that the English past tense, categorisation and categorisation with exceptions are closely related to each other whereas auto and arbitrary associations are closely related to each other. Moreover, heritability can potentially play a useful role as an indicator of task relatedness. Lastly, the approach proposed in this paper always results in good generalisation performance in all tasks and the average performance on all tasks keeps improving with generations. This clearly indicates that the results obtained using our proposed approach are due to the cumulative effects of intrinsic parameters and training datasets rather than chance.

VI. CONCLUSION

Transfer learning approaches face four main challenges: to successfully perform transfer in case of heterogeneous tasks; the lack of a generalised mechanism of determining task relatedness; avoiding negative transfer; and to imitate more closely learning as it happens in human beings – taking into account both structure and environment where the system is placed. In this work, we proposed a behavioural genetics based transfer approach. We simulated population studies, using artificial neural networks as computational models capable of learning various heterogeneous tasks, i.e. tasks with different features. Our work addressed these four open questions.

The proposed approach enabled a population of ANNs to learn five tasks which vary with respect to their characteristics such as features, degree of similarity between input-output patterns, the presence of structure or regularity in mappings and overall complexity. We simulated the effects of genetic influences via variations in the neuro-computational parameters of the ANNs and the effects of environmental influences via a filter applied to the training set based on socio-economic-status values. Focusing on heritability, we showed how the direction of change in heritability can be used as an indicator of task relatedness. The experimental results denote that if the change in heritability for different tasks move in the same direction, this implies that same set of intrinsic parameters are required for learning and thus transfer within those tasks would be successful.

We reported performance results for all five tasks over four generations. They illustrate that using the outlined method heterogeneous tasks were learned successfully; with good performance in all cases and compared our approach with random networks and literal transfer. The model also captured the emerging relatedness amongst tasks with English past tense, categorisation and categorisation with exceptions belonging to one group of related tasks, and auto association and arbitrary associations belonging to another group of related problems. Despite distinctions emerging with respect to relatedness amongst tasks, shared knowledge (genome and filtered training set) helps learning in all cases. These results are of course only preliminary. More in depth experimentation and analysis are required to establish long-term trends and emerging behaviour as the populations evolve.

REFERENCES

[1] Pan S.J., Yang, Q.: A survey on Transfer Learning. In IEEE Transactions on Knowledge and Data Engineering. 22, 10, 1345-1359 (2010).

[2] Silver, D.L. & Mercer, R.E.: The Parallel Transfer of Task Knowldege Using Dynamic Learning Rates Based on Measure of Relatedness, Connection Science, 8,: 2, page numbers (1996)

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

[4] Caruana, R., "Multitask Learning." Learning to Learn, S. Thrun and L. Pratt (Eds.), Kluwer Academic Publishers, (1997).

[5] Torrey, L., & Shavlik, J.: Transfer learning. Handbook of Research on Machine Learning Applications. IGI Global, 3, 17-35. (2009).

[6] Plomin, R.: The role of inheritance. In behavior. Science, 248, 183-188. (1990).

[7] Plomin, R.: Behavioural genetics in the 21st century. International Journal of Behavioural Development, 24, 30-34. (2000).

[8] Thomas, M. S. C., Ronald, A., & Forrester, N. A.: Modelling the mechanisms underlying population variability across development: Simulating genetic and environmental effects on cognition. DNL Tech report 2009-1(2009).

[9] Thomas, M. S. C., Forrester, N. A., & Ronald, A.: Modeling Socioeconomic Status Effects on Language Development. Developmental Psychology. Epub ahead of print 1 April (2013).

[10] Plomin, R., & Spinath, F. M.: Intelligence: Genetics, genes, and genomics. Journal of Personality and Social Psychology, 86, 112-129. (2004).

[11] Plomin, R., DeFries, J. C., & McClearn, G. E.: Behavioral Genetics: A primer. San Francisco: Worth and Freeman. (2008).

[12] Plomin, R. & Kovas, Y. "Generalist genes and learning disabilities.". Psychol Bull 131 (4): 592–617. (2005).

[13] Plomin, R., & Caspi, A. :Behavioral genetics and personality. *Handbook of personality: Theory and research*, *2*, 251-276. (1990).

[14] Pratt,L.Y.: Non-Literal Transfer among Neural Network Learners. Technical report number MCS-92-04, to appear as chapter in: Neural Networks: Theory and Applications II, R.J. Mammone and Y.Y.Zeevi,eds.,1993