Learning of Past Tense using the Connectionist approach

By: Abhishek Kr. Mall
Varun Sharma

Introduction

- What is a connectionist approach to language learning??
- In a connectionist approach, there are a set of input nodes and a set of output nodes which are connected to each other by weighted connections.
- There may be hidden layers in between also.

- Initially the weights defined are arbitrary.
- They get adjusted along with the learning process to give the correct output.
- Various learning algorithm for this approach are available.
- Connectionist approach is contrary to the dominant symbolic approach because it does not work on rule governed principles and their exceptions.

- Advantage in favor of this kind of approach is that it is said to model the brain more closely than the symbolic approach.

THE SIMULATION

Simulation by J. L. McClelland and D. E. Rumelhart

- One of the most famous simulation in the connectionist domain of learning.
- Their simulation showed the learning of past tense in the same way as a child learns past tense.
- They obtained the same U-shaped graph showing rote learning, overgeneralization and recovery from overgeneralization.

- Rumelhart and McClelland adopted the Wickelgren model to represent the verbs which included Wickelphones and Wickelfeatures.
- It almost permits a differentiation of all the root forms of english and their past tenses.

OUR MODEL

Model 1

- Initially we started with 26 nodes for each letter in input, hidden and output layer.
- For each letter the node was assigned a value of 0.5 and it was incremented by 0.5 on each representation.
- The obvious fault with this model was that words having same letter more than twice could not be represented and order could not be maintained.

- The model gave fair results on regular and irregular verbs if trained separately, but failed to give any conclusion if both categories were used simultaneously.
- The training method used was Bayesian back propagation. The number of iterations were 100.

Model 2

- Next we used a model which included 260 nodes , each of the 26 nodes had only one node among them activated in the order of the letter in the verb. This model could be used for words having less than equal to 10 letters.
- It had the advantage of letter ordering in it.
- The disadvantage was that the matrix of 260X number of words was very scarce and therefore pattern recognition would have been difficult and the errors would be large.

- But no conclusions could be drawn as the demand on computational capacity exceeded and we got:

  OUT OF MEMORY!!!!

Model 3 – The Final Model

- Finally we used the same classification of phonemes as used by Rumelhart and McClelland.
- In our model we can represent each phoneme using 6 bits.
- Our model is :-

		Front		Middle		Back	
		V/L	U/S	V/L	U/S	V/L	U/S
Interrupted Cons.	stop	b	p	d	t	g	k
	nasal	m	-	n	-	N	-
Continuous Cons.	fric.	v/D	f/T	z	s	Z/j	S/C
	liq/SV	w/l	-	r	-	y	-
Vowel	high	E	i	O	^	U	u
	low	A	e	I	a/@	W	*/o

- Since we have 2 different categories, each category has 3 different classification and each classification has 2 more types.
- This makes it 6 on each side and so we decided to use 3 bits for representing each of the two major categories.
- Thus each phoneme could be represented with 6 bits , 3 each from the two categories

- The bits 000000 represented nothing and was used to show no activations of the node.
- And the rest of the representation is as shown: -

stop	001
nasal	010
fric.	011
Liq/SV	100
high	101
low	110

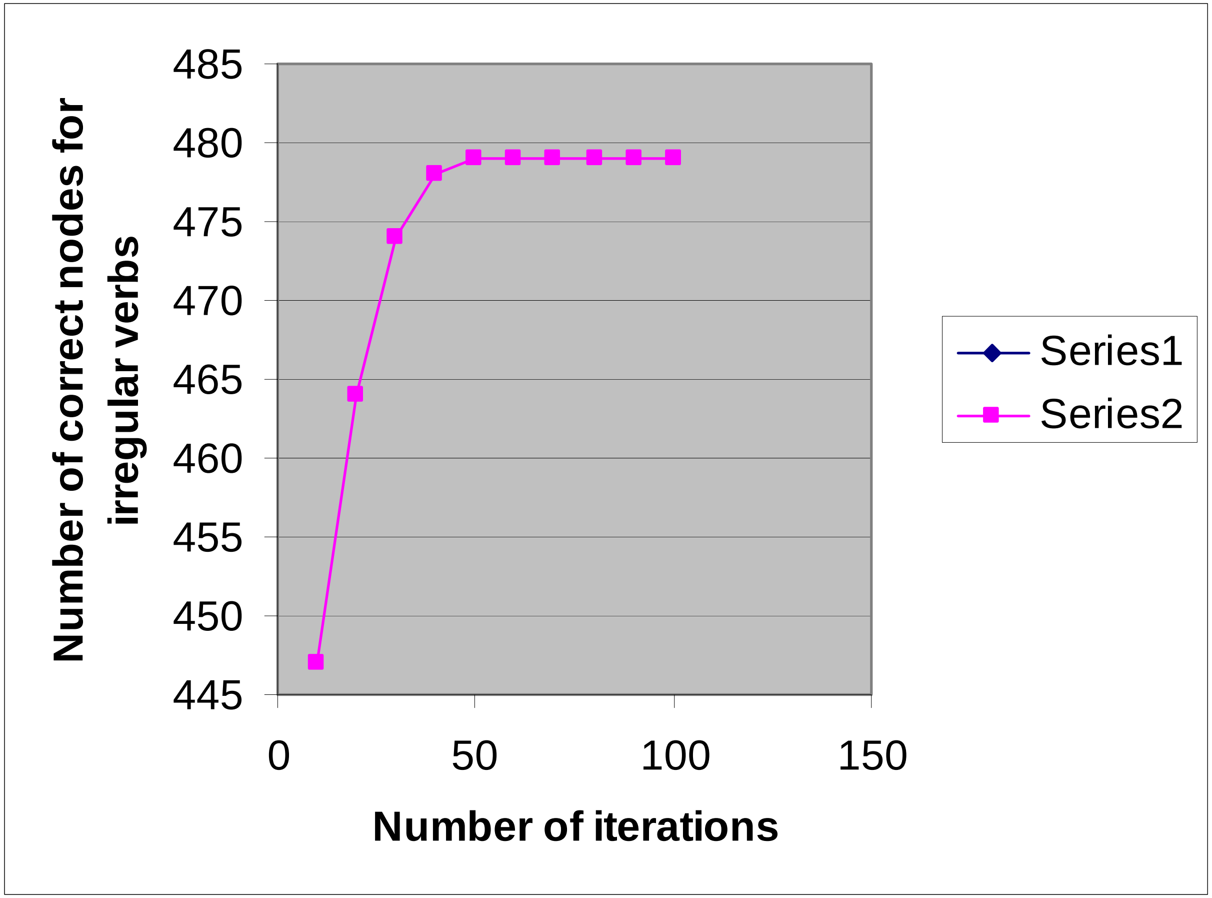
F v/l	001
F u/s	010
M v/l	011
M v/l	100
B v/l	100
B u/s	110

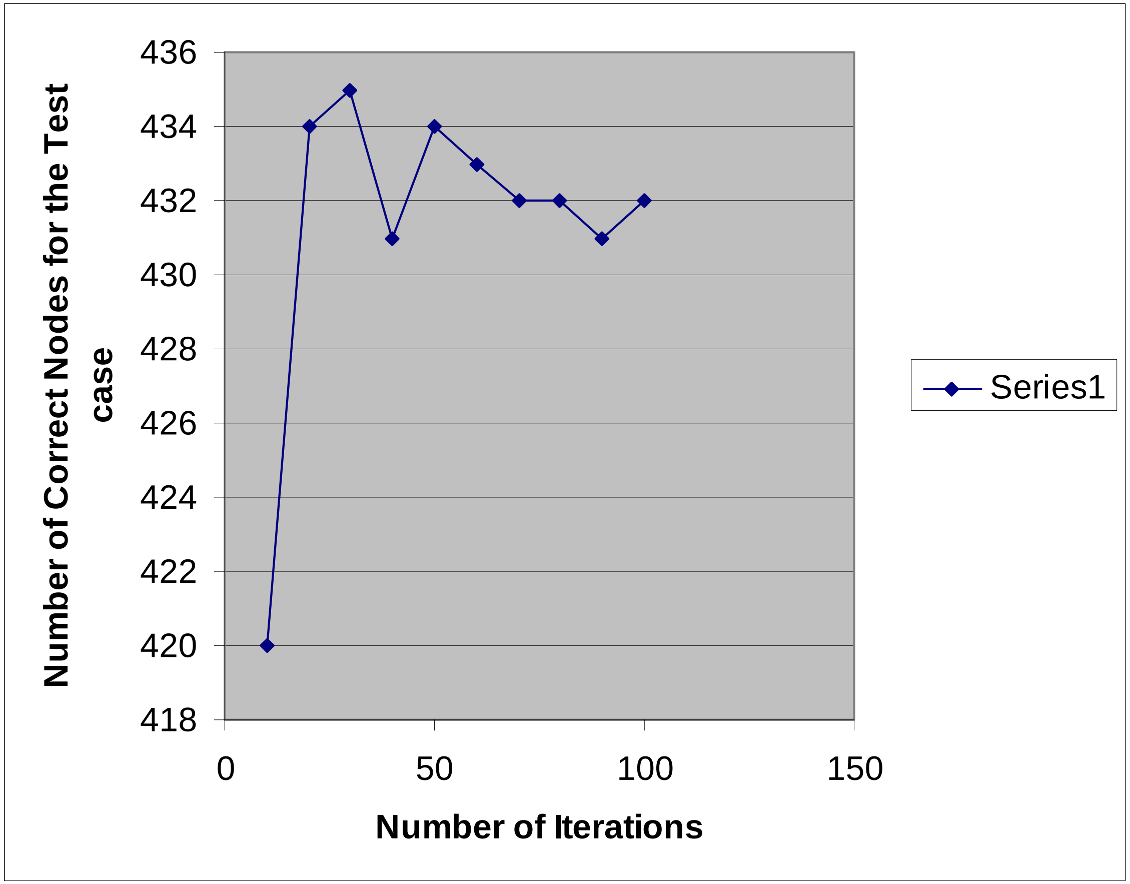
- We restricted the number of phonemes in a word to 8.
- This makes the array size for input to be 48Xnumber of words.
- We have used a Bayesian learning along with the Back propagation algorithm.
- We used 100 epochs and at an interval of 10 we calculated certain fixed parameters as shown afterwards.

Verb Data Used

- Our input training file included both regular and irregular words 100 in total.
- We made two test files one which had the 10 irregular verbs already taught to the network.
- The second one included 10 totally new verbs both regular and irregular.

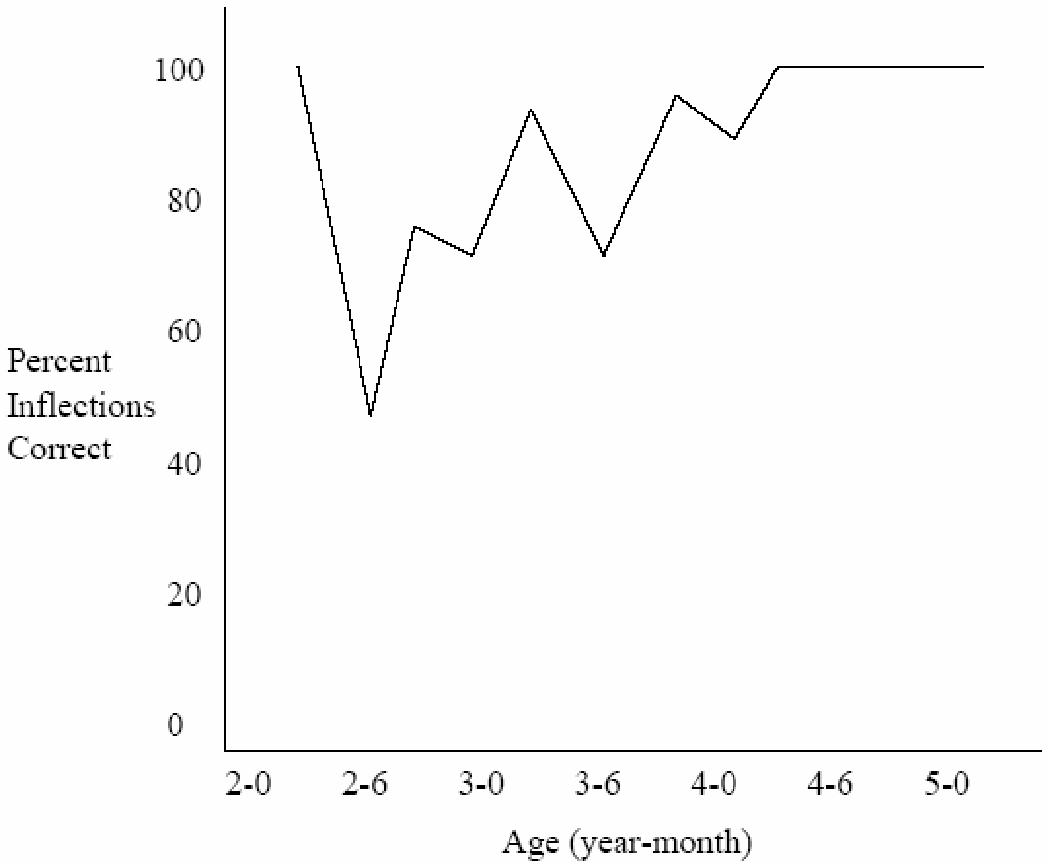
- In both the cases we drew a graph between the number of correct nodes that were activated for each of the two tests versus the number of iterations.
- The graphs we obtained were:-


485
480
475
470
465
460
455
450
445
0
50
100
150
Number of iterations
Number of correct nodes for irregular verbs
Series1
Series2



Number of Correct Nodes for the Test case
436
434
432
430
428
426
424
422
420
418
0
50
100
150
Number of Iterations
Series1

Reference Graph



Conclusion

- Our simulation has correctly learnt the verbs and their past tenses given as input.
- For the case when the input data had new verbs the results were found to be correct up to 89%.
- From the 2nd graph we can see that at first it over generalizes and then it is partially recovering which approximates the characteristic of child learning.

Disadvantages of our model

- The word length is limited which deviates from the fact that children learn words of even larger lengths.
- Our error (SSE) reaches to a critical point after sometime and does not tend to minimize the error further.

THANK YOU