Joseph Peña
Psych 209
Winter 2017

Psych 209 Final Project:
Modeling learning past tense verb inflection in English

Introduction

This project addresses the topic of regularization errors in past tense verb formation in English and is based on previous work by Kim Plunkett and Virginia Marchman on modeling learning past tense verb inflection in English using connectionist nets, *From rote learning to system building: acquiring verb morphology in children and connectionist nets* (1993).

Language acquisition by children has long been a topic that has interested me, as I've been fascinated both by the success of such a difficult task as language acquisition for children and by the bizarreness of utterances made by children of various ages.

This project serves as a launch point for the enrichment of my own understanding of neural networks—how they work and the computational situation for which they can be implemented—by implementing my own neural network, with the scaffolding of previous work, on a psychological, computational problem that interests me.

In Plunkett and Marchman (1993), a feed forward back propagation network (FFBP) is implemented to explore how incremental quantitative and structural changes to the verb vocabulary during training can give rise to qualitative shifts in network organization—specifically, they are looking for conditions that give rise to a U-shaped learning pattern, which is observed in children. This pattern consists of three stages: 1) generally all verbs are inflected correctly (by rote memorization), 2) then regularization errors occur (by competition between how different subclasses of verbs are treated), and 3) generally all verbs are inflected correctly again (by having found a working solution to this competition issue).

The exact purpose of my project will be to explore myself one specific aspect of this study: the effect of the network's vocabulary size on how the network handles novel, phonologically legal, indeterminate verb stems. (Indeterminate verb stems possess no defining characteristics about their verb class.) Per Plunkett and Marchman, the expected result is that as the vocabulary size of the network increases, the tendency to adopt suffixation (regularization) as the default mapping strategy also increases. Additionally, I will explore another of Plunkett and Marchman's findings: that additional training at a certain vocabulary size does not increase the rate of suffixation.

Method

Using the PDPyflow software, I implemented my own FFBP network. The network has 57 input units, 76 hidden units, and 95 output units. My training data comes from previous work by colleagues of Jay McClelland. In the data phonemes are represented by a 19 bit code. (Phonemes are mapped to bits based on factors such as place of articulation.) The network learns to map three-phoneme verb stems to five-phoneme past tense verbs. Inputs are oriented such that the third phoneme right before the verb ending is always in the same place. The network has a learning rate of 0.1 and momentum of 0. The hidden layer and output layer connections are randomly initialized in the range [-1, 1] and use the sigmoid activation function.

Following Plunkett and Marchman, the training data consists of 500 verb stems, which includes the following distribution of verb subclasses: 2 arbitrary verbs (e.g. go, went), 458 regular verbs, 20 identity verbs (e.g. hit, hit), and 20 vowel change verbs.

Plunkett and Marchman used only artificial, phonologically legal verb stems; however, given the availability of data, the accessibility of the data, and the time need to process the data to construct the vocabulary I used a combination of real English verbs and artificial verbs.

In my training data, the two arbitrary verbs are artificial. This was done to follow Plunkett and Marchman's guidelines that all verbs have three-phoneme stems and that there be no relation between verb stem and inflected form for arbitrary verbs. In addition, six identity verbs are artificial. This was done to meet the constraints that the number of identity verbs be twenty and that verb stems be three phonemes long. These verbs are a contraction of identity verbs with longer than three phonemes stems such that they still end in the /t/ or /d/ phoneme, as necessary for this subclass.

One additional note, Plunkett and Marchman, having created all of their verbs artificially, chose four subclasses for all vowel change verbs to fit into. Such a distribution of subclasses doesn't actually occur naturally in English. Given this, I chose to include verbs from the seven largest subclasses (out of 23 subclasses present) in the training data I had in order to create enough regularity of patterns in the vowel change verbs so that my network may succeed in mapping these verbs. These subclasses are (in my data's phonological code) aI – I, @U – u, I - V, aI - @U, i – e, I - &, and eI – U. Each subclass has two to four members.

Review of training set

Regular verbs (458)
Act Adore Cap Sheet Leak Miss Weave Piece Wait Claw Arch Move Turn Loose Love Flee Rib Rain Oust Lord … … …
Arbitrary verbs (2)
tiz → kige (k-aI-g) kurb → pal
Identity verbs (20)
Beat Bet Bid Cut Hit Hurt Let Put Rid Set Shed Shut Wed Wet Cot* Cat* Wit* Burt* Pit* Ped* [*constraint-fitted forms of identity verbs: Cost, Cast, Quit, Burst, Spit, and Spread]

Vowel change verbs (20)
Bite Blow Dig Feed Grow Hide Lead Light Meet Read Ride Ring Rise Shake Sing Sit Take Throw Win Write

The training procedure is based on Plunkett and Marchman's vocabulary expansion schedule. First, an initial training set is constructed following this distribution of subclasses: 2 arbitrary verbs, 10 regular verbs, 4 identity verbs, and 4 vowel change verbs. Verbs are chosen at random during construction. In addition, I followed Plunkett and Marchman's method for taking word frequency into account. The arbitrary verbs are each in the initial training set 15 times, and each of the other verbs 5 times.

This initial training set is trained on the network until "perfect". I've defined perfect as have an error less than 0.1 for each bit. Then an expansion schedule is carried out: every five epochs, a new verb is added to the training set. This new verb is chosen from an 80% regular, 20% not regular distribution. These new verbs are added with a frequency of three. This continues until the training set contains 100 verbs. After this, every epoch a new verb is added to the training set. This new verb is chosen from the same regular/not regular distribution and is added with a frequency of one. This continues until the training set contains all 500 verbs.

The network is tested on a list of 50 novel, artificial, indeterminate verb stems. These were adapted from data from McClelland's colleagues. Phonologically legal regular past tense inflections were created for these verbs. The distribution of regular past verb phonemes is 55% /d/ (28), 29% /t/ (14), and 16% /ed/ (8). This distribution matches the distribution found in the regular verbs in the total training set. In testing, an error of 0.1 for each bit is acceptable, as before with the initial training set.

Review of training and test set statistics

Vowel change (20)	Identity (20)	Regular (458)	Regular – Test (50)
I - V (2) eI - U (2) aI – I (3) @U - u (3) aI - @U (3) I - & (3) i - e (4)	ends in /d/ (15) ends in /t/ (15)	/d/ (253) - 55% /t/ (133) - 29% /ed/ (72) - 16%	/d/ (28) - 55% /t/ (14) - 29% /ed/ (8) - 16%

Review of statistics for each training stage

	Stage 1: Initial training	Stage 2: New verb every 5 epochs, frequency of 3	Stage 3: New verb every 1 epoch, frequency of 1
Number of verbs in vocabulary	20	100	500
	2 arbitrary (10%) 10 regular (50%)	2 arbitrary (2%) 78 regular (78%)	2 arbitrary (0.4%) 458 regular (91.6%)

	4 identity (20%) 4 vowel change (20%)	11 identity (11%) 9 vowel change (9%)	20 identity (4%) 20 vowel change (4%)
Number of tokens in vocabulary	120	360	760
	30 arbitrary (25%) 50 regular (41.7%) 20 identity (16.7%) 20 vowel change (16.7%)	30 arbitrary (8.3%) 254 regular (70.6%) 41 identity (11.4%) 35 vowel change (9.7%)	30 arbitrary (4.2%) 634 regular (83.4%) 50 identity (6.6%) 46 vowel change (6.1%)

Review of statistics of regular verbs for each training stage

	Stage 1: Initial training	Stage 2: New verb every 5 epochs, frequency of 3	Stage 3: New verb every 1 epoch, frequency of 1
Number of verbs in vocabulary	10 regular	78 regular	458 regular
	6 /d/ (60%) 4 /t/ (40%) 0 /ed/ (0%)	41 /d/ (52.6%) 25 /t/ (32.1%) 12 /ed/ (15.4%)	253 /d/ (55.2%) 133 /t/ (29%) 72 /ed/ (15.7%)
Number of tokens in vocabulary	50 regular	254 regular	634 regular
	30 /d/ (60%) 20 /t/ (40%) 0 /ed/ (0%)	135 /d/ (53.1%) 83 /t/ (32.7%) 36 /ed/ (14.2%)	347 /d/ (54.7%) 191 /t/ (30.1%) 96 /ed/ (15.1%)

Results

<u>Initial results</u>

	Epoch	Total error	**% verbs suffixed**	% /d/ *correctly* suffixed*	% /t/ *correctly* suffixed*	% /ed/ *correctly* suffixed*
after initial training	500	400.827	**18**	77.8	n/a	n/a
after vocab=100	900	156.421	**24**	100	33.3	0
after vocab=500	1258	67.75	**62**	100	70	0

* of the /x/ verbs suffixed at all, % of them that are suffixed correctly

<u>Averaged results</u>

These are averaged results over a sample of ten trials.

	% verbs suffixed	% /d/ *correctly* suffixed	% /t/ *correctly* suffixed	% /ed/ *correctly* suffixed
after initial training	**16.6**	70.5	85.2	n/a
after vocab=100	**26.2**	99	37	n/a - 0
after vocab=500	**57**	100	63.9	0

As expected, that the tendency to adopt suffixation as the default mapping strategy indeed increases as the vocabulary size increases.

<u>Halting vocab growth + Additional training</u>

Stopping at vocab=50: Averaged results from ten trials

	after initial training	after vocab=50	500 epochs later
% verbs suffixed	17	19.2	22.2

Stage 1 to Stage 2 increase (400 total epochs of training and vocab growth) = 2.2%

- Rate of change per epoch = 0.0055 %/epoch

Increase in suffixation after 500 additional training epochs = 3%

- Rate of change per epoch = 0.0050 %/epoch

Stopping at vocab=100: Averaged results from ten trials

	after initial training	after vocab=100	500 epochs later
% verbs suffixed	15.4	27.8	30.4

Stage 1 to Stage 2 increase (400 total epochs of training and vocab growth) = 12.4%

- Rate of change per epoch = 0.031 %/epoch

Increase in suffixation after 500 additional training epochs = 2.6%

- Rate of change per epoch = 0.0052 %/epoch

Analysis

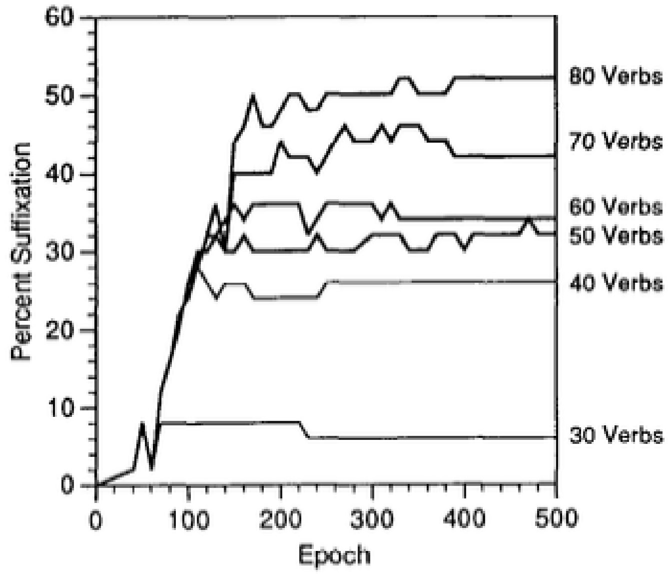
First, I'll start with why we see an increase in suffixation, touching on the particulars of each subclass of regular verbs. Then, I'll focus on the how halting vocabulary growth with additional training affects the rate of suffixation, a finding from Plunkett and Marchman that I found interesting.

Why do we see an increase in overall suffixation (regardless of correctness)? Simply, there are many more regular verbs in the vocabulary and tokens thereof in the training set that any other subclass of verbs. It's important to note that, while suffixation increases, in none of the trials on any test were all of the /d/ verbs, the most common (55.2%) regular verb subclass, suffixed. This is a result of the weight of the other subclasses given their influence in the initial training set, which is trained until "perfect." The number of tokens for regular /d/ verbs jumps from 30 in Stage 1 to 135 to 347 by Stage 3. Compared to the combined tokens of all other subclasses – 70 in Stage 1 to 106 to 126 by Stage 3 – we see that the other subclasses have more of an influence at the beginning but are quickly overcome in influence by /d/ regular verbs in the next two stages. Moreover, even given the many more epochs of training the Stage 1 verbs that were trained until "perfect" have, the new verbs added, which aren't trained to perfection, exert more of a force on mapping strategies than the initial training set, indicating that a more natural process of learning from the world, which involves an accumulation of exposure to quite varied different kinds of verbs, is more important for learning the strategy of regularizing novel past tense verbs than training until "perfect" on a closed training set, as we typically see with neural networks.

Why do we see a decrease and then an increase in correct suffixation for /t/ ending regular verbs? In Stage 1 there is an average of 85.2% of /t/ verbs that are suffixed being suffixed correctly, and in Stage 2 this drops to 37% and goes back up to 63.9% in Stage 3. For Stage 1, there were 6 trials with 1 suffixation, 2 trials with 2 suffixations, 1 trial with 3, and 1 trial with zero. Given the statistics of the training set – 40% of the regular verbs were /t/ verbs – the network gains fairly good exposure to how to map appropriate verbs to the /t/ ending, and this can explain why at the beginning sometimes a /t/ verb would be mapped correctly. However, during the next two training stages, the network receives a very different proportion of exposure to regular verbs: in Stage 2 and 3 the proportion of tokens of regular verbs is about 55%/30%/15% (d/t/ed). This creates competition for strategies for mapping regular verb stems to their appropriate endings. This is corroborated by /t/ verbs very often being mapped to /d/ in Stage 2 and 3, reflected by the 37% correctly suffixed in Stage 2 and 63.9% suffixed in Stage 3. The increase then from Stage 2 to Stage 3 can be explained as the network finding a partial solution to this competition issue.

Why can't the network correctly suffix /ed/ ending regular verbs? No /ed/ verbs in any trial were suffixed at all in Stage 1. Only in 3 trials in Stage 2 were any /ed/ verbs suffixed, though none correctly, and while some /ed/ verbs were suffixed in every trial in Stage 3, none were correctly suffixed here either. Firstly, there are so few /ed/ verbs in the whole vocabulary and by chance none in the initial training set. /ed/ regular verbs make up only 14.4% of the entire vocabulary. The percent of tokens that are /ed/ grows from 0% to 14.2% to 15.1%. It is interesting to note though that many /ed/ verbs in Stage 1 were mapped to the "empty" ending. This can be explained by the influence of identity verbs in the training set and onward, as the stems for identity verbs end in either the phoneme /d/ or /t/ just like the stems of /ed/ verbs do. The poor performance of /ed/ verbs here is most personally interesting for me, as I've experienced both a 3-year-old native speaker and an 8-year-old English learner mysteriously use, what I then perceived as, the present tense of a verb instead of the past tense (i.e. drop the /ed/ ending completely).

Now I will focus on how halting vocabulary growth with additional training affects the rate of suffixation. In Plunkett and Marchman (1993), they found if vocabulary growth were halted at any vocabulary size and then they continued training for many for epochs that the rate of suffixation would remain about steady.



Plunkett and Marchman (1993)

My data is inconclusive in showing this same effect. With more programming ability, I could have attained more conclusive data, but nevertheless I'll proceed with what I have. In stopping vocabulary growth at 50 and then training for an additional 500 epochs, the rate of change in suffixation per epoch decreased from 0.0055 %/epoch to 0.0050 %/epoch. In stopping vocabulary growth at 100 and continuing training, the rate of change in suffixation per epoch decreased from 0.031 %/epoch to 0.0052 %/epoch. While in both cases, there is a decrease and it could be the case that the rate of suffixation is steadying out, many more data points at different numbers of additional training epochs and different sizes of vocabularies would be need for more conclusive results.

Discussion

Taking Plunkett and Marchman's data (above) offers a starting point for this discussion. This steady rate of suffixation even given more training seems to be more evidence for what I said previously about the influence of exposure to more verbs (adding verbs to the training set) compared to the influence of more training on a closed set. It's interesting that even though in each case of vocabulary size the training set will eventually reach "perfection" by more training, this doesn't affect the rate of suffixation for novel verbs. There is evidence that accumulated exposure to many kinds of verbs is more important for adopting the regularization strategy for novel verbs is more important than exposure to a consistent set of verbs.

This conclusion matches well a picture of the world. It is important to note that the language children are exposed to, while certainly coming from a certain probability distribution for classes of words in a given language (as taken into account with the creation of the initial training set and the vocabulary expansion schedule, per Plunkett and Marchman), is also accumulated quite haphazardly, and this particular accumulation of language exposure is also (more?) important in forming our understandings of how our language works – which can be demonstrated by the fact that native speakers of a language tend to inflect novel, artificial verbs regularly.

Limitations of this work include the formation of the total verb vocabulary, in which all verbs had to begin with three phonemes and vowel changing verbs were poorly represented compared to their naturally occurring distributions. The training until "perfect" of the initial training set is also only a very rough approximation of an infant's exposure to verbs – not to mention that no other aspects of language that may affect how we come to understand verbs are taken into account in this model. My own limitation in programming and scripting affected the quality of data I attained in my results from halting vocabulary growth. And lastly, it's worth mentioning, that modeling learning says nothing definitively about how the brain does these computation – it provides us only an example of how such a computation can work.