Instructions for ACL 2019 Proceedings

Anonymous ACL submission

Abstract

Several features of human language can be described through rules, and natively speaking children of languages are cognizant of these rules, even without formal instruction. Prior research has attempted to model such learning in children using connectionist neural networks for paradigms such as the English past tense. We investigate how a gradient-descent neural network could acquire the rules for declining Russian nouns into the genitive case for both singular and plural nouns in a similar fashion. We also test whether two different NNs for the singular and plural numbers can model the learning of these rules in children better than a single NN by comparing the models' performance to that of native Russian-speaking children.

1 Credits

This paper is an adaptation of the paper *On Learning the Past Tenses of English Verbs* by D. E. Rumelhart and J. L. McClelland.

2 Introduction

Many features of natural languages, such as verb conjugation, noun case declension, and agglutination behave in ways that can be described by sets of rules. Being cognizant of these rules is an important aspect of language acquisition in children, because awareness of these rules allows them to utilize newly acquired vocabulary to express new concepts in the same manner they may have used in a previous utterance. For example, the most common ending used to express the past tense in English is -ed, as in "I waited at home". Awareness of this rule would allow a child to additionally express "I hunted at the park" after learning the verb "hunt" using the same ending.

However, it is also often the case that these rules have exceptions or other irregularities. In English, ablaut is a source of several irregular verbs' past tense forms. For example, the past tense of "run" is "ran", and a child aware of the basic past tense rule may incorrectly express "I runned to the park" as a result, a phenomenon known as overgeneralizing. This phemoenon is also important for learning, as the corrections the child receives improve their understanding of the past tense, ultimately allowing them to generalize the past tense for more verbs with greater accuracy.

Prior research has attempted to create neural networks to function as language acquisition devices (LAD) in order to model how children acquire the English past tense. Most notably, Rumelhart and McClelland (1986) made use of a perceptron model to model the learning of English past tense in three stages based on prior research into how children acquire it, and obtained learning results that very closely align with children's learning. We therefore investigate whether the acquisition of the genitive case in Russian would yield similar results.

2.1 Russian noun case morphology

Grammatical cases are a system of marking nouns and noun modifiers in order to indicate the noun (modifier)'s intended grammatical function in a sentence. While historically English had grammatical cases, they survive in modern English only through pronouns. Every pronoun in English has three different forms representing three different cases. These are:

1. Nominative: Marks the subject of a sentence: *I* help him.

2. Accusative: Marks the indirect object of a sentence: He helps *me*.

3. Genitive: Marks the possessor of a noun: *My* book is great.

Russian, being a member of the Slavic branch of the Indo-European languages, has a fully functioning grammatical case system which is used for all nouns unlike English. Nouns and noun modifiers in Russian inflect for case by changing the ending of the word, and the exact manner by which an ending changes depends on the ending, case, and number of the noun. Table 1 shows the different possible endings of the Russian nominative and genitive cases in the singular and plural.

There are, of course, various irregularities in forming the genitive case for certain nouns. Some nouns have suppletive roots, found only in the plural form: for example, the noun *čelovjek* (person) has the regular singular genitive form *čelovjeka*. However, in the nominative plural, it becomes *ljudji*, and in the genitive plural *ljudjej*, using an entirely different noun stem for the plural cases.

The reason we chose the genitive case to study is because its formation, much like the English past tense, is fairly regular in the singular; in a majority of cases, the genitive form can be simply derived from the nominative by modifying the noun's ending based on the table. We also chose it because the genitive plural, in particular, tends to be more irregular. It is often the aspect of Russian that is most difficult for learners of the language to comprehend, so we wanted to investigate how a gradient-descent neural network would handle acquiring the genitive case in singular and plural forms.

We also chose the genitive because it is a case that (with the exception of indeclinable nouns, which we ignored for this paper) always changes the noun in some way. This is contrasted with the accusative case, which can either not change the noun, use the genitive form to represent the accusative, or use its own unique endings. Whichever scenario applies depends on not just the ending of the noun, but also its gender and animacy.

3 Featural representation of Russian phonemes

Rumelhart and McClelland (1986) made use of a scheme to encode English phoneme units, known as Wickelphones and Wickelfeatures, described in Wickelgren (1969). Wickelphones are phoneme units representing each phone in a word as triples, which consist of the phoneme itself, its predecessor to the left, and its successor on the right. Wickelfeatures are a means of representing Wickelphones as distributed patterns of activation by capturing one feature of the central phoneme, one feature of the predecessor, and one feature of the successor.

	Singular	**Plural**
Nominative	-a, -ja, -ija	-I\, -i, -ii
Genitive	-I\, -i, -ii	∅, -’, -ij
Nominative	∅, -’/j, -ij	-I\, -i, -ii, -je
Genitive	-a, -ja, -ija	-ov, -jej/jev, -ijev
Nominative	-o, -je	-a, -ja
Genitive	-a, -ja	∅, -j, -jej
Nominative	-’	-i
Genitive	-i	-jej
Nominative	-mja	-mjena
Genitive	-mjeni	-mjon

Table 1: Nominative and genitive endings for nouns with different endings, written with X-SAMPA. Note that the apostrophe ’ is a palatalization character, corresponding to ь in Cyrillic.

The feature coding scheme we used to capture Russian phonemes did not differ heavily from Rumelhart and McClelland's scheme. However, in order to accomodate Russian's richer consonant inventory, we had to add an extra dimension to the scheme to account for plain vs. palatalized consonants, the latter of which are in abundance in Russian phonology.

In addition to accomodating Russian phonology for Wickelfeatures, we had to make some decisions regards to what exact sounds are in Russian. Specifically, the status of the close central unrounded vowel /ɨ/ (written in Cyrillic as ы) as a separate vowel morpheme in Russian is a subject of heavy debate. Some linguists believe in a five-vowel analysis, meaning that the vowel /ɨ/ is in complementary distribution with the close front unrounded vowel /i/ (written in Cyrillic as и), such that the vowel phoneme /i/ occurs after soft (i.e. palatalized) consonants, whereas the vowel phoneme /ɨ/ occurs after hard (i.e. plain) consonants. Other linguists believe in a six-vowel analysis, asserting /ɨ/ as a separate phoneme.

Another set of phonemes with disputed status in Russian are palatalized velar consonants (i.e. /k^j/, /g^j/, and /x^j/) as well as the palatalized voiceless alveolar sibilant affricate /$\widehat{ts}^j$/. In most cases, the velar consonants become soft when followed by front vowels (except in the case of a word boundary between the consonant and vowel), and the affricate is generally always hard. It is only in certain loanwords and foreign names that the consonants may become palatalized in other contexts.

For the sake of our experiment, however, we have chosen to follow the five-vowel analysis and not count the extra palatalized consonants as Russian phonemes. The phonemic status of the vowel /ɨ/ is marginal, only occuring isolated in the verb ыкать (to pronounce the sound ы) and in borrowed names of places in Russia. By following the five-vowel analysis, we can account for one less vowel, because /ɨ/ and /i/ are otherwise in complementary distribution in all other environments. Likewise, because the velar consonants are normally only palatalized following front vowels, there is no need to account for such redundant information.

Table 2 shows the scheme that we used to categorize Russian phonemes on five simple dimensions.

4 Experimental setup

To run our experiment, we first obtained a Russian word list, filtered the word data, then preprocessed the data into a format that can be used by a neural network model. We then trained our neural network models based on our word data and evaluated the results. The procedure for these steps is outlined in the subsections below.

4.1 Collecting Russian nouns and declensions

In order to train the neural network, we needed a list of Russian nouns and their genitive singular and plural forms. For this, we scraped words from Wiktionary. Wiktionary is an online, multilingual, and free project to create a dictionary for all natural languages. These dictionary entries often contain additional information about words in languages, such as verb conjugations, example sentences, synonyms and antonyms, and most importantly, noun declensions.

Every Russian noun with an entry on Wiktionary is accessible through the category page **Category:Russian Nouns**. Using Wiktionary's MediaWiki API, we gathered every noun in this category in batches of 500 (due to API limitations). For each batch, we checked to make sure that the word only contained letters in the modern Russian Cyrillic alphabet, because Wiktionary also has entries for Russian words containing characters from different alphabets (e.g. QR-код, meaning "QR code") as well as alternate spellings of words containing archaic letters unused since the Russian spelling reform of 1918.

After performing this initial pruning, we then queried the API for each word's entry in Wiktionary and attempted to extract its genitive singular and plural forms from the declension table available in the entry. If either the singular or plural genitive forms were missing from the table, the noun is discarded. This can happen if a word is an indeclinable loanword or only exists in either the singular or plural forms. We then save the word, its genitive singular and plural forms into a wordlist, which will be consulted during the preprocessing and training phases.

4.2 Noun data pre-processing

Before the noun data can be used as input to train the neural network, it must first be converted into a representation usable by the model.

The first transformation that takes place is converting the noun forms' Russian Cyrillic representation into a phonemic representation in X-SAMPA. Because the Russian alphabet is much more phonemically consistent than English, meaning it is possible to derive a word's phonemic representation by iterating through its Cyrillic representation one letter at a time.

However, phonemic representations are not specific enough for the neural network. As previously referenced, Rumelhart and McClelland (1986) refer to Wickelgren (1969) to devise a featurization system using Wickelphones and Wickelfeatures. In particular, each Wickelphone has a set of at most 25 Wickelfeatures that its presence activates, which are derived from the phonological features of all three phonemes present in the Wickelphone.

For all phonemes other than the special word-boundary marker (which has a single unique 'boundary' feature), each phoneme has five features that correspond to the dimensions outlined in Table 2. For each feature in the central phoneme of the Wickelphone, at most five Wickelfeatures are generated: each feature in the left context phoneme, the selected feature of the central phoneme, and each feature in the right context. An example of this for the Wickelphone $_{\mathbf{j}}\mathbf{a}_{\mathbf{z}}$ is present in Table 3. Computationally, this algorithm corresponds well to a series of vector repetitions and a matrix transposition.

		Place											
		Front				Middle				Back			
		Voiceless		Voiced		Voiceless		Voiced		Voiceless		Voiced	
		H	S	H	S	H	S	H	S	H	S	H	S
Interrupted	Stop	p	p’	b	b’	t	t’	d	d’	k	-	g	-
	Nasal	-	-	m	m’	-	-	n	n’	-	-	-	-
Cont.Cons.	Fricative	f	f’	v	v’	s	s’	z	z’	s`	t_s\	z`	-
	Liq/Sem.	-	-	l	l’	t_s	-	r	r’	x	-	-	j
Vowel	Close	-	-	i	-	-	-	-	-	-	-	u	-
	Open	-	-	e	-	-	-	a	-	-	-	o	-

Table 2: Categorization of Russian phonemes on five dimensions. All phonemes are written using X-SAMPA.

Each of these activated Wickelfeatures are then “blurred”: for a Wickelfeature $\langle f_1, f_2, f_3 \rangle$, all Wickelfeatures of the form $\langle *, f_2, f_3 \rangle$ and $\langle f_1, f_2, * \rangle$ have a random chance of being activated as well. Rumelhart and McClelland (1986) claim that this improves the neural network’s ability to generalize between phonological features.

However, to avoid computational blowup, Wickelfeatures with different dimensions for their left and right features (other than the special word boundary feature) are filtered out of the computation. This brings the total number of possible Wickelfeatures down from 2028 to only 660.

Activations across Wickelfeatures are thus represented by a boolean 660-vector, where a 1 in index k corresponds to Wickelfeature k being activated. A word’s total activation is all the Wickelfeatures activated by the word’s composite Wickelphones. This is easily implemented with a vector sum operation.

4.3 Neural network

The basis of our neural network is a gradient-descent neural network powered by the *PyTorch* library. It uses a single linear transformation layer with the input size being the number of Wickelfeatures that can be activated for, and the output size being either 1 or 2 times the number of Wickelfeatures that can be activated for, depending on whether the model is one of the two smaller models learning individual genitive cases or the larger model learning both genitive cases.

During training, the network’s input is the activation vector for the nominative singular form of a noun, and the expected output is the activation vector for that noun’s genitive form (singular or plural, depending on the network being trained). However, the larger model attempting to learn both cases at once simply has the singular and plural genitive activation vectors concatenated together, with the intent being that the model will eventually output two “words” at once.

#	Left Context	Central Phoneme	Right Context
1	Continuous	Vowel	Continuous
2	Liquid	Vowel	Fricative
3	Back	Vowel	Middle
4	Voiced	Vowel	Voiced
5	Soft	Vowel	Hard
6	Continuous	Open	Continuous
7	Liquid	Open	Fricative
8	Back	Open	Middle
9	Voiced	Open	Voiced
10	Soft	Open	Hard
11	Continuous	Middle	Continuous
12	Liquid	Middle	Fricative
13	Back	Middle	Middle
14	Voiced	Middle	Voiced
15	Soft	Middle	Hard
16	Continuous	Voiced	Continuous
17	Liquid	Voiced	Fricative
18	Back	Voiced	Middle
19	Voiced	Voiced	Voiced
20	Soft	Voiced	Hard
21	Continuous	Hard	Continuous
22	Liquid	Hard	Fricative
23	Back	Hard	Middle
24	Voiced	Hard	Voiced
25	Soft	Hard	Hard

Table 3: The Wickelfeatures activated by the Wickelphone $_{\mathbf{j}}\mathbf{a}_{\mathbf{z}}$.

4.4 Decoding model output

Described in the Appendix of their paper, Rumelhart and McClelland (1986) use what they term a “binding network” to decode the feature activation vectors outputted by the learning network into a full phonological representation. Because the activation vector encoding scheme has no representation of the temporal dimension of a word (i.e., it does not directly encode *where* features and source phonemes occur in the source word), a separate decoder is necessary. The binding network does not learn its weights, but rather each weight is determined by the strength of its previous output “in time”, proportional to the probability of the corresponding feature being activated (i.e., the network output).

The network attempts to model how each Wickelphone contributes to the Wickelfeatures present in the activation vector, with the idea being “to find a set of output features [Wickelphones] that accounts for as many as possible of the output features while minimizing the number of input features [Wickelfeatures] accounted for by more than one output feature. Thus, we want each of the output features to *compete* for input features” (Rumelhart and McClelland, 1986, pp. 269).

5 Results

Unfortunately, our results were decidedly inconclusive. The neural networks did successfully learn during training, with a final loss of approximately 0.1 (down from 0.16 after 100 epochs). However, the decoder network did not work as hoped, and only output two Wickelphones when given output from any of the neural networks. For example, when given the network predicted activations for the input язык (X-SAMPA **jazik**), the decoder only ever output the Wickelphones ${}_{\#}\mathbf{j}_{\mathbf{a}}$ and ${}_{\mathbf{j}}\mathbf{a}_{\#}$; for the input жена (X-SAMPA **z’ena**), the decoder only returned ${}_{\#}\mathbf{z'}_{\mathbf{e}}$ and ${}_{\mathbf{z'}}\mathbf{e}_{\#}$.

5.1 Discussion

These results seem to suggest that a featural representation of words that properly preserves temporal information is necessary to get meaningful output from these models. It is also possible that the specification for the binding network in Rumelhart and McClelland (1986) was overly general, and thus some nuance in implementation may have been missed.

6 Conclusion

References

David E. Rumelhart and James L. McClelland. 1986. On learning the past-tenses of english verbs: implicit rules or parallel distributed processing.

Wayne A. Wickelgren. 1969. *Context-Sensitive Coding in Speech Recognition, Articulation and Development*, pages 85–96. Springer Berlin Heidelberg, Berlin, Heidelberg.

A Appendices

Appendices are material that can be read, and include lemmas, formulas, proofs, and tables that are not critical to the reading and understanding of the paper. Appendices should be **uploaded as supplementary material** when submitting the paper for review. Upon acceptance, the appendices come after the references, as shown here. Use `\appendix` before any appendix section to switch the section numbering over to letters.