An Extension of Elman Network and a Generalized Simple Recurrent Network

Shin-ichi Asakawa (asakawa@ieee.org)
Tokyo Womens' Christian University,
Centre for Information Sciences, 2-6-1 Zempukuji,
Suginami 1678585 Tokyo, JAPAN

Abstract

Although the Elman network is so powerful that it can deal with a variety of language processings, there exist some short comings about its ability. For example, the original Elman net cannot always deal with a long distance dependency appropriately, which is a number agreement between nouns and verbs with many relative pronouns in a sentence. This limitation might cause from the constraints of its structure of the context and the hidden layer, which can preserve only one time previous state of the network. Here, we propose an extension of the Elman network and a generalized simple recurrent neural network. The proposed networks could preserve the n-th generations of inner states. When the model processed the corpus consisted of many relative pronouns with multi-center embeddings structure, it could deal with the long distance number agreement adequately. These models can be regarded as a natural extension of the simple recurrent neural networks in order to deal with complex structures of language.

Keywords: Simple recurrent neural network; memory capacity; number agreement; long distance dependency.

The sturcture of the Elman network

The networks proposed by Jordan and Bishop(1996) and Elman (1990, 1991) were simple recurrent neural networks(SRN). These networks have an input layer in order to deal with the current input signals, and have a context layer which maintain past states. The contents dealt in the hidden layer at time t depend on both current inputs and past states at time $t-1$. Therefore, the context layer in SRN can maintain the whole history of all the past inputs. As a result, the state of the network at time t depends on both current inputs and the set of all the history of past inputs. Since there exists a computational limitation (a limitation of precision), SRN can only represent finite regions in a problem space. Therefore, a number of discussions revealed that SRN cannot overcome the limitation of finite state machines. It is known that any attempts to let SRN learn context dependent grammar have, more or less, a limitation that SRN cannot reach the same result as generalized finite state machines. However, here we will try to show possible solutions to tackle this problem.

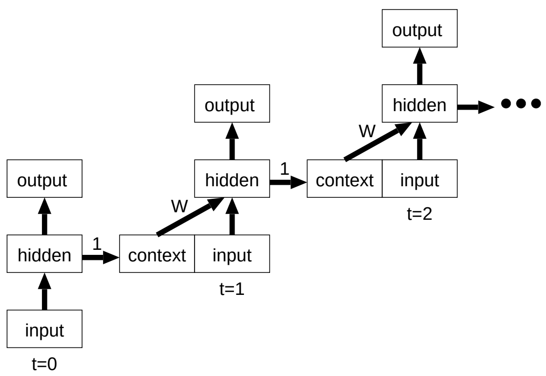

Figure 1: Time extension of SRN

The device to deal with the language which has center embeddings must have memory stores to maintain complex time sequential information. A number of experimental studies revealed that SRN could recognize and learn regular languages (eg. Giles et al.,1992). The SRNs are simple but powerful in order to deal with the context dependent grammar by the ability of time development. However, there is a limitation in SRN. It cannot deal with complex structure such as a long distance dependency like a number agreement between nouns and verbs in sentences with many relative pronouns (a nested center embedding structure). Although Elman found that the Elman network can deal the number agreement in center embedded sentences, it might be impossible to deal with many relative pronouns in multi-center embedded sentences by the limitation of the capacity or the state of the hidden and the context layers.

Extensions of context layers

The possible ways to overcome this problem are to extend volumes and contents of context layers in the network. There are three possibilities to extend context layers: (1) the extension of the number of units in the hidden and the context layers and (2) the extension of generations of context layers (Fig.2), and (3)the extension of contexts from various sources(Fig.3). The extension of the number of units in the context layer is a simple solution for the network to find the solution of the complex time information such as long distance dependencies. But the network cannot necessarily get the precise information which occurred the past. Although the extension of the number of the units in the context layer can enrich the information in the hidden and context layers, this extension might not result in a realistic solution of the long distance dependency. On the other hand, the extension of the number of generations in context layers is a direct way to deal with past n-th generations of contents in the inner state of the network. The n-th generations of context layers can receive exactly the same information of $n-1$-th generations of context layers. The extension of the number of generations in context layers might deal with this problem (multi-center embedded structure) adequately. Since the n-th context layer can receive the contexts of the $n-1$-th context layer directly, the network can deal with the information which occurred long time ago (n-th generation ago).

We can get more general framework than Elman's and Jor-

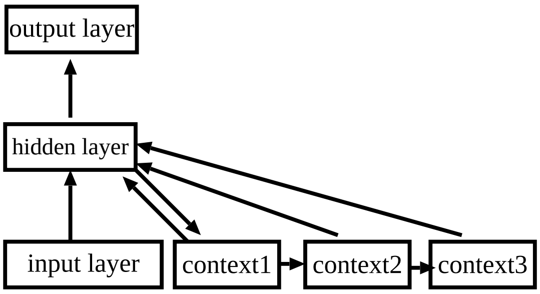

Figure 2: The extension of number of generations in context layers

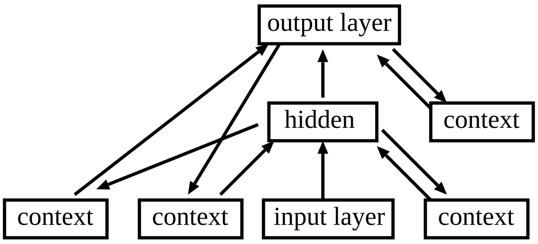

Figure 3: The extension from various sources

dan's networks by adding all the connections both hidden and output layers. Taking advantage of this framework, we can get an optimal architecture to solve general problems.

Numerical experiment

In order to confirm the ability of the networks with the extension of the number of generations in context layers and with the extension of multi-sources, a numerical experiment was performed. The sentences were generated in accordance with the grammar which was almost the same as the one which Elman (1991) used (see Table 1). Total 25 words included EOS (End Of Sentence) consisted of the 25 input and output units.

The normal Elman network which has 20 hidden and context units and the extended Elman network with 5 context layers were compared. The output of the normal Elman network with 20 hidden and context units is for example:
Mary(girl) feeds(who) lives walks. Mary(boy) who(walks). Mary lives. Mary(cats) who(hear) . The words in the parentheses indicate the correct answers. The number agreement between nouns and verbs preserved in shorter sentences. For example, *Mary(girl) who lives(chases) cat sees.* However, in case of longer sentences which have many relative pronouns, there was a tendency to show incorrect words, which means that the error words did not consist with the parts of speeches as the correct words. On the other hand, the extended Elman network with the 5 generations context layers could deal with the long distance dependency, for example *Mary(boy) who hears cat(Mary) sees Mary.*

Table 1: The grammar used in the experiment.

S	→	NP VP "."
NP	→	PropN | N | N RC
VP	→	V (NP)
RC	→	who V | who V NP | who V (NP)
N	→	boy | girl | cat | dog |
		boys | girls | cats | dogs
PropN	→	Mary | John
V	→	chase | feed | walk | live
		chases | feeds | walks | lives
		see | hear | sees | hears

Additional restrictions:
1. number agreement between N & V within clause, and (where appropriate) between head N & subordinate V
2. verb arguments:
chase, feed → require a direct object
see, hear → optionally allow a direct object
walk, live → preclude a direct object
(observed also for head/verb relations in relative clauses)

Conclusion

Although there was no significant difference in the sense of the quantity in total performance among the normal Elman network, the extended Elman network and a generalized SRN, there was a quality difference in errors in the sentences with many relative pronouns. The extended Elman network with multi-context layers could process sentences with many relative pronouns properly in the meaning that it could deal with long distance dependencies with multi-center embeddings. This might mean a potential ability of the extension in the number of generations in context layers.

In formal language theory, context dependent grammar is in general undecidable. The extended Elman network with multi-context layers and a generalized SRN proposed hear could be one of the possible candidates to deal with complex problems. These models can be regarded as a natural extension of the simple recurrent neural network with multi-memory storage. It could also be analogous with a human model of language information processing.

References

Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, *14*, 179–211.

Elman, J. L. (1991). Distributed representations, simple recurrent networks, and grammatical structure. *Machine Learning*, 195–225.

Giles, C. L., Miller, C. B., Chen, D., Chen, H. H., Sun, G. Z., & Lee, Y. C. (1992). Learning and extracting finite state automata with second-order recurrent neural networks. *Neural Computation*, *4*, 393–405.

Jordan, M. I., & Bishop, C. M. (1996). Neural networks. In *Crc handbook of computer science.* MIT Press.