Connectionist Context Processing for Phrase Filtering

Stefan Wermter[1]
Department of Computer Science
University of Hamburg
2000 Hamburg 50, Federal Republic of Germany

Email: wermter@nats4.informatik.uni-hamburg.de

Abstract

This paper describes connectionist recurrent plausibility networks that are examined for learning a scanning understanding of a large number of "real-world" phrases. These plausibility networks encode incremental context of phrases in recurrent connections and therefore allow for filtering phrases according to different context classes. Since input representations are automatically acquired as significance vectors these plausibility networks support the portability across different domains. Furthermore, since underlying regularities for context assignment are learned from an unrestricted corpus of phrases, these networks support adaptability and robust processing independent of the underlying grammatical constructions.

1 Introduction

In the past, there have been several approaches for text filtering in the fields of symbolic natural language processing and information retrieval. These approaches classify phrases, sentences, or complete texts according to various context classes. However, approaches from symbolic natural language processing usually do not use any learning and symbolic matching rules have to be manually encoded (e.g., [2]). Furthermore, the learning of a graded incremental context representation of natural language phrases is not considered. On the other hand, approaches for text filtering from information retrieval are fast, simple, and robust but accuracy of the retrieval classifications as measured with recall and precision is very low for typical classification tasks [6]. One major reason for this low performance is the lack of sequential context knowledge. In contrast, recurrent connectionist representations have the potential to learn underlying sequential regularities for context classification and to represent plausible knowledge. Previous recurrent networks showed the ability to express sequentiality using recurrent connections [3] [1] [4]. However, so far the examined networks were developed for restricted domains, often with artificially generated training material and it is not clear to what extent such models scale up in a "real-world" environment. Furthermore, since phrases contain much less predictive semantic knowledge associated with verb forms than complete sentences, a scanning understanding of phrases requires an approach that is based on a stronger bottom-up semantic memory.

In our general approach of a scanning understanding we pursue the representation of syntactic, semantic, and contextual knowledge for the interpretation of phrases. This approach is examined in

[1]This research was supported in part by the Federal Secretary for Research and Technology under contract #01IV101A0.

a general architecture SCAN based on a modular, plausibility-based, constraint-based, and learned use of hybrid connectionist representations for phrases [7]. In this paper we will particularly concentrate on connectionist learning of graded incremental representations for contextual phrase analysis. Learning a contextual phrase context provides important knowledge for instance for the tasks lexical disambiguation, structural disambiguation and semantic interpretation [8]. Furthermore, we explore connectionist learning techniques to increase adaptability, portability, robustness of phrase analysis and to scale up context processing to a substantial number of "real-world" phrases.

2 Training of Recurrent Plausibility Networks

Recurrent plausibility networks represent the incremental context for a context assignment in reduced internal representations (hidden layers). Figure 1 shows the general structure of a recurrent plausibility network. A feedforward network is extended with distributed recurrent delays at hidden layers. The input to a hidden layer L_x is constrained by the underlying feedforward layer L_{x-1} as well as the incremental context at different preceding time steps $t-1$ to $t-t_x$ where t_x is the maximum number of recurrent time steps for layer L_x. The connections between two adjacent layers with n and m units are fully connected $n : m$ connections. This architecture is trained with an extension of backpropagation for recurrent plausibility networks [5] [7]. The architecture extends simple recurrent networks and ensures that internal representations of the context can be used at different time steps so that not only the directly preceding context but also initial partial context is available later.

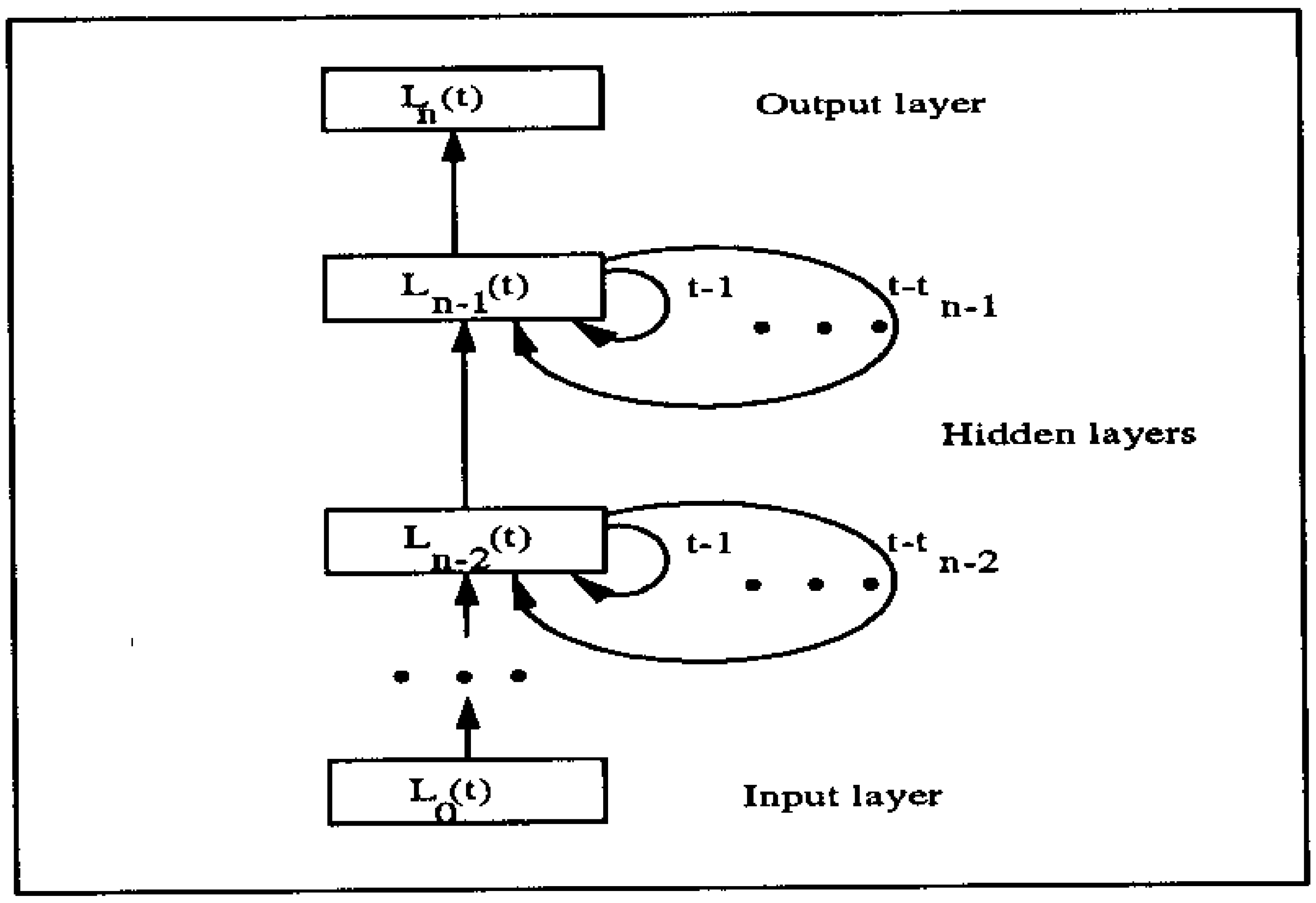

Figure 1: General structure of a recurrent plausibility network

For representing phrases and determining its semantic context class we used a recurrent plausibility network since its partially recurrent connections consider the incremental context of an unrestricted phrase. Each phrase is presented to the network as a sequence of single significance vectors together with their context class. Each significance vector consists of 10 units that represent the normalized frequency of occurrences of this word within 10 context classes. These significance vectors have been computed based on a library corpus of title phrases with 6110 phrases

(30206 words). The context classes were: theology, history, law, mathematics, chemistry, computer science, electrical engineering, geology, art, and music. (A typical example phrase would be "Measurement of high-speed signals in solid state devices" from the electrical engineering context.) During training, the hidden layer develops an internal representation of the preceding words in a phrase (word-context relationships). In our experiments we tested context layers (hidden layers) with three to twenty units. The output layer contains one unit for each of the ten context classes. The output unit which represented the desired context class was set to 1 and all other output units were set to 0.

To examine this architecture for complete unrestricted phrases, we selected a training set of 1000 unrestricted phrases from 10 different context classes of the library corpus. Then the generalization behavior of the network was tested with another 1000 phrases whose representations as sequences of significance vectors were different from representations in the training set. We performed three runs each for networks between three and twenty units for one hidden layer. A configuration with 10 units performed best. In the beginning of the training, when the error of the assigned context classes was still high, we used a small learning rate (0.000001 for 200 epochs) and a weight change momentum of 0.9. The small learning rate led to smaller weight changes in the beginning of the training and prevented the network from changing the weights too fast in different directions. Then, after the network had learned some important regularities, the learning rate was increased to 0.00001 and the network was trained for another 200 epochs. Increasing the learning rate at that point sped up the learning compared to keeping the learning rate constant.

As an alternative to using compete phrases in a recurrent plausibility network, a scanning understanding might benefit from a processing strategy that emphasizes significant domain-dependent words. Therefore, we removed insignificant words from the unrestricted title phrases that belong to the categories determiners, prepositions, conjunctions, and pronouns and that occur more than four times in our complete corpus. These experiments were conducted with 1000 training titles, 1000 test titles, the same learning rates, weight change momentum, number of epochs, runs, and number of hidden units as for the complete unrestricted titles. Similarly as for the complete titles, the network with ten hidden units performed best.

3 Results of Semantic Context Classification

Text filtering, text extraction, and other related tasks usually use two terms, recall and precision, to describe retrieval tasks. In our context, *recall* describes the percentage of phrases from class X which have been assigned to class X. *Precision* describes the percentage of phrases assigned to class X which really belong to class X.

Evaluation	Recall	Precision
complete training phrases	98.8%	98.1%
complete test phrases	96.8%	95.0%
training phrases without insignificant words	98.5%	98.5%
test phrases without insignificant words	95.4%	95.8%

Table 1: Recall and precision for classifying 2000 phrases into 10 context classes

Table 1 shows that results on the training set are slightly better than on the test set for complete phrases (98.8% versus 96.8% for recall, 98.1% versus 95.0% for precision) and for condensed phrases without insignificant words (98.5% versus 95.4% for recall, 98.5% versus 95.8% for precision). Furthermore, using complete phrases compared to condensed phrases provides slightly better results for recall (98.8% versus 98.5% for training, 96.8% versus 95.4% for testing). On the other hand, the use of condensed phrases provides slightly better results for precision (98.5% versus 98.1% for training, 95.8% versus 95.0% for testing). That is, the elimination of context-ambiguous insignificant words leads to a weaker recall rate but better precision. However, in general we receive very good recall and precision results, that is at least 95% for both, recall and precision.

4 Conclusion

We have described connectionist recurrent plausibility networks that allow for representing the incremental preceding context in phrases. This context representation was tested by classifying a substantial number of 2000 "real-world" title phrases from a library classification. Since the underlying input representation was acquired automatically, these techniques can be transported to various other domains. Furthermore, we primarily relied on the contextual-semantic knowledge of sequences of significance vectors, which leads to very robust processing with no restrictions with respect to the grammatical structure of the covered phrases. Thus, this work shows that plausibility networks can *learn* a scanning understanding of classifying and filtering unrestricted natural language phrases. Therefore, these recurrent connectionist plausibility networks are potentially important for scaling up classification tasks in natural language processing and for increasing adaptability and transportability across different domains.

References

[1] J. L. Elman. Structured representations and connectionist models. In *Proceedings of the Eleventh Conference of the Cognitive Science Society*, pages 17–25, Ann Arbor, MI, 1989.

[2] P. J. Hayes, L. E. Knecht, and M. J. Cellio. A news story categorization system. In *Proceedings of the Second Conference on Applied Natural Language Processing*, pages 9–17, Austin, TX, 1988.

[3] M. I. Jordan. Attractor dynamics and parallelism in a connectionist sequential machine. In *Proceedings of the Eighth Conference of the Cognitive Science Society*, pages 531–546, Amherst, MA, 1986.

[4] J. B. Pollack. Recursive distributed representations. *Artificial Intelligence*, 46:77–105, 1990.

[5] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. Learning internal representations by error propagation. In D. E. Rumelhart and J. L. McClelland, editors, *Parallel Distributed Processing*, volume Vol. 1, pages 318–362. MIT Press, Cambridge, MA, 1986.

[6] G. Salton. *Automatic Text Processing*. Addison Wesley, New York, 1989.

[7] S. Wermter. A hybrid connectionist approach for a scanning understanding of natural language phrases. Technical Report Doctoral thesis, University of Hamburg, Hamburg, FRG, 1993, forthcoming.

[8] S. Wermter and W. G. Lehnert. A hybrid symbolic/connectionist model for noun phrase understanding. *Connection Science*, 1 (3):255–272, 1989.