Recurrent Neural Learning for Helpdesk Call Routing *

Sheila Garfield and Stefan Wermter

University of Sunderland, Sunderland, SR6 0DD, United Kingdom
{stefan.wermter,sheila.garfield}@sunderland.ac.uk
http://www.his.sunderland.ac.uk/

Abstract. In the past, recurrent networks have been used mainly in neurocognitive or psycholinguistically oriented approaches of language processing. Here we examine recurrent neural networks for their potential in a difficult spoken language classification task. This paper describes an approach to learning classification of recorded operator assistance telephone utterances. We explore simple recurrent networks using a large, unique telecommunication corpus of spontaneous spoken language. Performance of the network indicates that a semantic SRN network is quite useful for learning classification of spontaneous spoken language in a robust manner, which may lead to their use in helpdesk call routing.

1 Introduction

Language is not only extremely complex and powerful but also ambiguous and potentially ill-formed [1]. Problems associated with recognition of this type of speech input can result in errors due to acoustics, speaking style, disfluencies, out-of-vocabulary words, parsing coverage or understanding gaps [5]. Spontaneous speech also includes artifacts such as filled pauses and partial words. Spoken dialogue systems must be able to deal with these as well as other discourse phenomena such as anaphora and ellipsis, and ungrammatical queries [5, 9].

In this paper we describe an approach to the classification of recorded operator assistance telephone utterances. In particular, we explore simple recurrent networks. We describe experiments in a real-world scenario utilising a large, unique corpus of spontaneous spoken language.

2 Description of the Helpdesk Corpus

Our task is to learn to classify real incoming telephone utterances into a set of service level classes. For this task a corpus from transcriptions of 4 000 recorded operator assistance telephone calls was used [2]. The utterances range from simple direct requests for services to more descriptive narrative requests for help as shown by the following examples:

* The authors thank Mark Farrell and David Attwater of BTexact Technologies for their helpful discussions.

1. "*could I um the er international directory enquiries please*"
2. "*can I have an early morning call please*"
3. "*could you possibly give me er a ring back please I just moved my phone I want to know if the bell's working alright*"

Examination of the utterances reveals that the callers use a wide range of language to express their problem, enquiry or to request assistance [3].

2.1 Call Transcription

The focus of the investigation was a corpus from transcriptions of the first utterances of callers to the operator service. Analysis of the utterances identified a number of service levels or call class categories, primary move types and request types [2]. The primary move is a subset of the first utterance and is like a dialogue act and gives an indication of which dialogue act is likely to follow the current utterance.

Four separate call sets of about 1 000 utterances each were used in this study. The call sets are split so that about 80% of utterances are used for training and approximately 20% of utterances used for testing. The average length of an utterance in the training set is 16.05 words and in the test set the average length of an utterance is 15.52 words. Each call class is represented in the training and test set. An illustrative example is given in Table 1, however not all call classes are shown. The part of the utterance identified as the primary move was used for both the training and test sets. At this stage some utterances were excluded from the training and test sets because they did not contain a primary move utterance.

Table 1. Breakdown of utterances in training and test sets from call set 1. Note: For illustration purposes not all classes are shown

917 utterances										
Total of 712 utterances in Training set										
Total of 205 utterances in Test set										
Categories:	class 1	class 2	class 3	class 4	class 5	class 6	class 7	class 8	class 9	class n
in train set:	261	11	41	3	85	32	6	16	28	...
in test set:	59	3	21	1	29	11	2	4	7	...

The number of call class categories in the task is 17 and call set 1 has an entropy of 3.2.

$$entropy = \sum_{i=1}^{17} P(c_i) log_2(P(c_i)) \tag{1}$$

2.2 Vectors for Semantic SRN Network

The experiments use a semantic vector representation of the words in a lexicon [8]. These vectors represent the frequency of a particular word occurring in a call class category and are independent of the number of examples observed in each category. The number of calls in a class can vary substantially. Therefore we *call normalize* the frequency of a word w in class c_i according to the number of calls in c_i (2). A *value* $v(w, c_i)$ is computed for each element of the semantic vector as the *normalized* frequency of occurrences of word w in semantic class c_i, divided by the *normalized* frequency of occurrences of word w in all classes. That is:

$$Norm.\ freq.\ of\ w\ in\ c_i = \frac{Freq.\ of\ w\ in\ c_i}{Number\ of\ calls\ in\ c_i} \tag{2}$$

where:

$$v(w, c_i) = \frac{Norm.\ freq.\ of\ w\ in\ c_i}{\sum_j Norm.\ freq.\ for\ w\ in\ c_j},\ j \in \{1, \cdots n\} \tag{3}$$

Each call class is represented in the semantic vector. An illustrative example is given in Table 2, however not all call classes are shown. As can be seen in the illustrative example, domain-independent words like 'can' and 'to' have fairly even distributions while domain-dependent words like 'check' and 'order' have more specific preferences.

Table 2. Example of semantic vectors. Note: For illustration purposes not all classes are shown. There are 17 classes used in this study

Word	Call Class									
	class 1	class 2	class 3	class 4	class 5	class 6	class 7	class 8	class 9	class n
CAN	0.03	0.06	0.09	0.07	0.05	0.08	0.04	0.04	0.04	...
YOU	0.07	0.01	0.01	0.05	0.09	0.10	0.05	0.06	0.05	...
JUST	0.18	0.00	0.00	0.00	0.03	0.03	0.00	0.07	0.00	...
CHECK	0.53	0.00	0.00	0.00	0.00	0.00	0.11	0.00	0.06	...
TO	0.08	0.06	0.01	0.05	0.02	0.01	0.05	0.08	0.13	...
SEE	0.36	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	...
IF	0.12	0.08	0.02	0.00	0.06	0.10	0.05	0.08	0.03	...
IT	0.19	0.00	0.00	0.06	0.04	0.06	0.01	0.04	0.09	...
OUT	0.10	0.00	0.00	0.00	0.02	0.12	0.00	0.12	0.07	...
OF	0.09	0.00	0.00	0.11	0.14	0.01	0.03	0.00	0.06	...
ORDER	0.24	0.00	0.54	0.00	0.00	0.00	0.00	0.00	0.00	...

3 Learning and Experiments

A semantic SRN network with input, output, hidden and context layers was used for the experiments. Supervised learning techniques were used for training [4, 7]. The input to a hidden layer L_n is constrained by the underlying layer L_{n-1} as well as the incremental context layer C_n. The activation of a unit $L_{ni}(t)$ at time t is computed on the basis of the weighted activation of the units in the previous layer $L_{(n-1)i}(t)$ and the units in the current context of this layer $C_{ni}(t)$ limited by the logistic function f.

$$L_{ni}(t) = f(\sum_{k} w_{ki} L_{(n-1)i}(t) + \sum_{l} w_{li} C_{ni}(t)) \tag{4}$$

This provides a simple form of recurrence that can be used to train networks to perform sequential tasks over time. Consequently, the output of the network not only depends on the input but also on the state of the network at the previous time step; events from the past can be retained and used in current computations. This allows the network to produce complex time-varying outputs in response to simple static input which is important when generating complex behaviour. As a result the addition of recurrent connections can improve the performance of a network and provide the facility for temporal processing.

3.1 Training Environment

In one epoch, or cycle of training through all training samples, the network is presented with all utterances from the training set and the weights are adjusted at the end of each utterance. The input layer has one input for each call class category. During training and test utterances are presented sequentially to the network one word at a time. Each input receives the value of $v(w, c_i)$, where c_i denotes the particular class which the input is associated with. Utterances are presented to the network as a sequence of word input and category output representations, one pair for each word. Each unit in the output layer corresponds to a particular call class category. At the beginning of each new sequence the context layers are cleared and initialised with 0 values. The output unit that represents the desired call class category is set to 1 and all other output units are set to 0. An utterance is defined as being *classified* to a particular call class category if at the end of the sequence the value of the output unit is higher than 0.5 for the required category. This output classification is used to compute the recall and precision values for each utterance. These values are also used to compute the recall and precision rates for each call class category as well as the overall rates for the training and test sets.

The network was trained for 1 000 epochs on the training transcribed utterances using a fixed momentum term and a changing learning rate. The initial learning rate was 0.01, this changed at 600 epochs to 0.006 and then again at 800 epochs to 0.001. The results for this series of experiments are shown in Table 3.

3.2 Recall, Precision and F-Score

The performance of the trained network in terms of recall, precision and F-score on the four call sets is shown in Table 3. Recall and precision are common evaluation metrics [6]. The F-score is a combination of the precision and recall rates and is a method for calculating a value without bias, that is, without favouring either recall or precision. There is a difference of 3.54% and 5.11% between the highest and the lowest test recall and precision rates respectively.

Table 3. Overall results for the semantic SRN network using semantic vectors

	Training Set			Test Set		
	Recall	Precision	F-Score	Recall	Precision	F-Score
Call Set 1:	85.53%	93.84%	89.49	79.02%	90.50%	84.37
Call Set 2:	84.26%	92.76%	88.31	75.48%	87.22%	80.93
Call Set 3:	87.06%	93.72%	90.27	76.00%	85.39%	80.42
Call Set 4:	85.47%	93.15%	89.14	76.38%	85.39%	80.63

4 Analysis of Neural Network Performance

The focus of this work is the classification of utterances to service levels using a semantic SRN network. In general, the recall and precision rates for the semantic SRN network are quite high given the number of service levels available against which each utterance can be classified and the ill-formed input. The semantic SRN network achieved an average test recall performance of over 76% of all utterances. This result is calculated based on the overall performance figures for the semantic SRN network shown in Table 3.

In other related work on text classification [8] news titles were used to classify a news story as one of 8 categories. A news title contains on average about 8 words. As a comparison, the average length of the first caller utterance is 16.44 words and is subject to more ambiguity and noise. On the other hand, the size of the vocabulary used in the text classification task was larger than that used for our classification of call class categories. The performance of the simple recurrent network is significant when this factor is taken into consideration because a larger vocabulary provides more opportunity for the network to learn and therefore generalise on unseen examples. While on an 8 category *text* classification task we reached about 90%, in this study presented in this paper here for a much more ill-formed *spoken language* classification task and 17 categories we reached above 75% (recall) and 85% (precision) for unseen examples.

We have compared our results also with a feedforward network without recurrent connections. Recurrent networks performed significantly better. The better performance of the SRN network shows that the network does make use of the

memory introduced by the context layer to improve both its recall and precision rates. This shows that the information stored in the context layer, which is passed back to the first hidden layer, does assist the network in assigning the correct category to an utterance.

5 Conclusions

In conclusion the main aim of this research is to identify indicators about useful semantic SRN architectures that can be developed in the context of a larger hybrid symbolic/neural system for helpdesk automation. A description has been given of a recurrent neural architecture, the underlying principles and an initial evaluation of the approach for classifying the service level of operator assistance telephone utterances. The main result from this work is that the performance of the SRN network is quite good when factors such as noise in the utterance and the number of classes are taken into consideration. This work makes a novel contribution to the field of robust learning classification using a large, unique corpus of spontaneous spoken language. From the perspective of connectionist networks it has been demonstrated that a connectionist network, in particular a semantic SRN network, can be used under *real-world* constraints for spoken language analysis.

6 Acknowledgments

This research has been partially supported by the University of Sunderland and BTexact Technologies under agreement ML846657.

References

[1] Abney, S.: Statistical Methods and Linguistics. In: Klavans, J., and Resnik, P. (eds): The Balancing Act. MIT Press, Cambridge, MA (1996)
[2] Durston, P.J., Kuo, J.J.K., et al.: OASIS Natural Language Call Steering Trial. Proceedings of Eurospeech, Vol **2**. (2001) 1323–1326
[3] Edgington, M., Attwater, D., Durston, P.J.: OASIS - A Framework for Spoken Language Call Steering. Proceedings of Eurospeech '99. (1999)
[4] Elman, J.L., Bates, E.A., Johnson, M.H., Karmiloff-Smith, A., Parisi, D., Plunkett, K.: Rethinking Innateness. MIT Press, Cambridge, MA (1996)
[5] Glass, J.R.: Challenges for Spoken Dialogue Systems. Proceedings of IEEE ASRU Workshop. Keystone, CO (1999)
[6] Salton, G., McGill, M.: Introduction to Modern Information Retrieval. McGraw Hill, New York (1983)
[7] Wermter, S.: Hybrid Connectionist Natural Language Processing. Chapman and Hall, Thomson International, London, UK (1995)
[8] Wermter, S., Panchev, C., Arevian, G.: Hybrid Neural Plausibility Networks for News Agents. Proceedings of the National Conference on Artificial Intelligence. Orlando, USA (1999) 93–98
[9] Clark, H.: Speaking in Time. Speech Communication **36** (2002) 5–13