2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)
October 11-14, 2020. Toronto, Canada

Learning Co-Occurrence of Laughter and Topics in Conversational Interactions

Kristiina Jokinen[1] , Junpei Zhong[2*]

Abstract—**This paper describes experiments to learn laughter co-occurrences with dialogue contributions. The dialogue data belongs to the special type of First Encounter Dialogues where the interlocutors meet each other for the first time and where laughter mainly functions as a sign of politeness or relief of embarrassment. The earlier studies have shown that there is a correlation between the speaker's utterance content (topic) and non-verbal communication (laughter and body movement) while in this paper we seek to learn the correlations via a neural model. The results show that there seems to be a weak correlation in our data.**

Index Terms—**component, formatting, style, styling, insert**

I. INTRODUCTION

As argued in [1], interactions in social robotics are situated, i.e. interactions take place in a dynamically changing world, with different physical settings and among participants with varying skills. This imposes further requirements for the interaction modelling as the situations in real world are more complex and cannot be fully specified in advance. Interactive robot systems which can talk with humans have been developed (e.g. [2], [3]) and the basic technology for building intelligent interactive agents is already matured to the level of commercial applications. However, the current interactive systems, have very straightforward interactions with the user and usually they do not take the user's non-verbal communication or affective state into account in the interaction modelling. In order to make the interactions more natural, it is important to add flexibility in the system's dialogue management strategy (cf. [4]). For instance, it is relevant to understand the participant's utterances and the utterance content (dialogue topics) as well as how the utterance content and the participant's affective state are correlated, so as to be able to provide situationally appropriate responses. From the point of view of social robotics, capability for such behaviour is crucial since it shows understanding and attention to the partner's needs. This is important in order to create a positive and acceptable atmosphere for the interactive situation, i.e., to increase usability and user satisfaction of the service application. For instance, if it is possible to distinguish the user's amused and benevolent state from a neutral state or somewhat irritated state, this could be used to generate a system reaction that aligns with the user's affective state: to provide a slightly humorous or light-hearted response to increase bonding and partnership, or a neutral explanation to continue on factual interaction, or even ask a direct query about the user's interest to show immediate reaction and care of the user's uncomfortable situation. Moreover, earlier research suggests that laughter and dialogue topics are correlated, so if it is possible to predict possible laughter occurrences in the dialogue on the basis of utterance content, this may be of valuable help in managing and structuring the dialogue in terms of dialogue topics.

In our previous work ([5], [6]), we used the First Encounter Dialogues and aligned laughter (as annotated in the speech files) with the participants' body movement (as automatically recognized from the videos), and the results conformed the original hypothesis that body movements are in synchrony with laughter. However, the conclusion also emphasised the complexity of the problem and that more systematic studies are needed with a larger data set. In this paper we continue with these studies but focus on the utterance content and its correlation with laughter occurrences using neural methodology.

[7], [8] studied the interlocutors' engagement and involvement in human-human conversations, and noticed that laughter and topic management are correlated: laughter as a social signal seems to structure topic changes and its timing conveys information about the underlying discourse structure. In particular, topic transition points usually have higher amounts of laughter than topic continuation points, i.e. when a topic is "exhausted" and a topic change occurs, also more laughing takes place, as if a sign of "relief". Moreover, when the temporal distance from the topic boundary increases, laughter becomes more likely to occur, and a significant change in the amount of laughter occurs at fifteen seconds around the topic changes.

Conventional topic modelling has mainly focussed on single modality environment on static documents, i.e. on large texts collected from archives, such as newspapers, journals, logs, on-line chats, etc. [9], [10]. Topic detection task is usually tackled as a clustering problem [11], involving feature selection and k-means oriented techniques, and graph-based methods for event detection and multimodal clustering in social media streams have also appeared. However, topic clustering often results in overlapping topics which do not always accord with human intuition in their semantics and temporal length. Although topics change and evolve in the course of the dialogue and it may be difficult to determine accurate topic changing points, the previous research seems to implicate that laughing that occurs within the conversation has an important role to play: ”exhausted” topics or topics which are embarrassing or difficult to converse over, are associated with laughter, which thus signals the participants' willingness to change the topic or stop talking about the particular topic.

* Corresponding author: zhong@junpei.eu

[1] Artificial Intelligence Research Center, National Institute of Advanced Industry Science and Technology, Tokyo, 135-0064, Japan

[2] Nottingham Trent University, Nottingham, NG11 9NS, United Kingdom

978-1-7281-8526-2/20/$31.00 ©2020 IEEE

In this paper, we thus focus on the utterance content and experiment with neural modelling approach and end-to-end dialogue techniques. We study how word2vec embeddings and the LSTM (Long Short-Term Memory) RNN work in predicting the laughter occurrences.

The contributions of the paper are:

- learning of neural model for dialogue topics;
- learning the correlation of topics and laughing;
- improving naturalness of dialogue systems with the help of laughter model;
- experimenting with the applicability of a small data set.

A. On Laughter

Laughter is a typical social behaviour, commonly associated with joking and humour [12], and it serves a broad range of interaction functions. It can be a sign of politeness and socially acceptable behaviour in casual situations where a generally benevolent and friendly relation is to be displayed, but it can also occur in connection to socially critical situations to reduce psychological tension or embarrassment. Conversation Analysis (e.g. [13]) has studied various regularities concerning the structural position and functions of laughing in social interactions, based on the work by [14] who talks about situated interactions and how the bursts of laughter can structure interactions by breaking the ordinary interactional frames. While laughing can signal positive feedback to create common understanding and rapport among the participants, it is also an acceptable way to disassociate oneself from the topic of the conversation.

Laughter is much studied in speech research from the acoustic point of view, for the purposes of emotion recognition or speech synthesizer ([15], [16], [16], [17], etc.) Classifications of laughter often distinguish free laughter from speech-laugh, i.e. laughter which is synchronous with speech. This distinction is also the basis for our classification, see discussion on the data annotation and the annotation tags in Tab. I and II.

II. Data and Method

A. Data Description

First encounter dialogues are a special type of dialogues where the participant meet for the first time, and in a short time are meant to learn to know each other more. The participants are bound by social conventions (e.g. politeness codes how to talk with strangers) but they are also fairly free to conduct conversations on any suitable topic and switch the topics as they find appropriate.

Our dialogue corpus consists of an Estonian video corpus of first encounter dialogues, collected in the MINT project [18]. The data consists of 23 dialogues with 23 participants (12 male and 11 female) who were native speakers of Estonian. Each participant had two encounters, but with a different partner. The participants were instructed to get to know their partner in a short conversation as they might do at a party or a reception. The participants were students and university employees with the age between 21 and 61 years. There are 8 female-female encounters, 7 female-male encounters, and 8 male-male encounters, each about 8 minutes long.

The university context makes interactions less formal and more on the level of collegial interaction; also teacher-student interactions are friendly and natural rather than impose a formal hierarchical relationship.

The corpus was transcribed and manually annotated using the ELAN annotation tool. The annotations concern the participants' head, hand and body movements as well as laughter occurrences. The laughter occurrences follow loosely the general laughter annotations used in speech studies (e.g. [15], [17], [19]). The occurrences are first divided into free laughs and speech-laughs, and further into subtypes which loosely relate to the speaker's affective state. The division between free and speech laughs is rather even: 57% of the laugh occurrences are free laughs. However, the different subtypes have unbalanced distribution as seen in I. This may reflect the fact that the first encounters are basically friendly and benevolent interactions among fairly confident young adults.

The total number of laughs is 530, average 4 occurs in a second.

B. Method description

To give a straightforward impression to obtain an assumption, we firstly visualise part of the dialogues in the dataset. The Fig. 1 demonstrates some examples in different coloured symbols about how the laughter occurs with the context of certain topics in the time-line of in this conversational interaction. The blue and red bar in the middle indicates the 1st and the 2nd person appearing in the conversation. And the symbols appearing on top or at the bottom of the bar correspond to different types of laughter in the dataset as shown in Tab. II.

From the visualisation samples, we make a hypothesis that there is a co-occurrence relation between the types of laughter and the semantics of the topics in the conversation. To discover such a relation with the laughter and the topics, a recurrent neural network model with the word embedding method is used to learn it. Although there are a lot of factors about the occurrence of laughter, for instance, the timing, the pause in-between, etc. For the sake of simplicity, in this work, we let the output of the model be the 12 types of the laughter as shown in Tab. II, and the input is the temporal sequence of the words in the conversation that occurs before and after the very laughter 3 with the window width w. The words are represented in the form of the word embedding learnt with a pre-trained model.

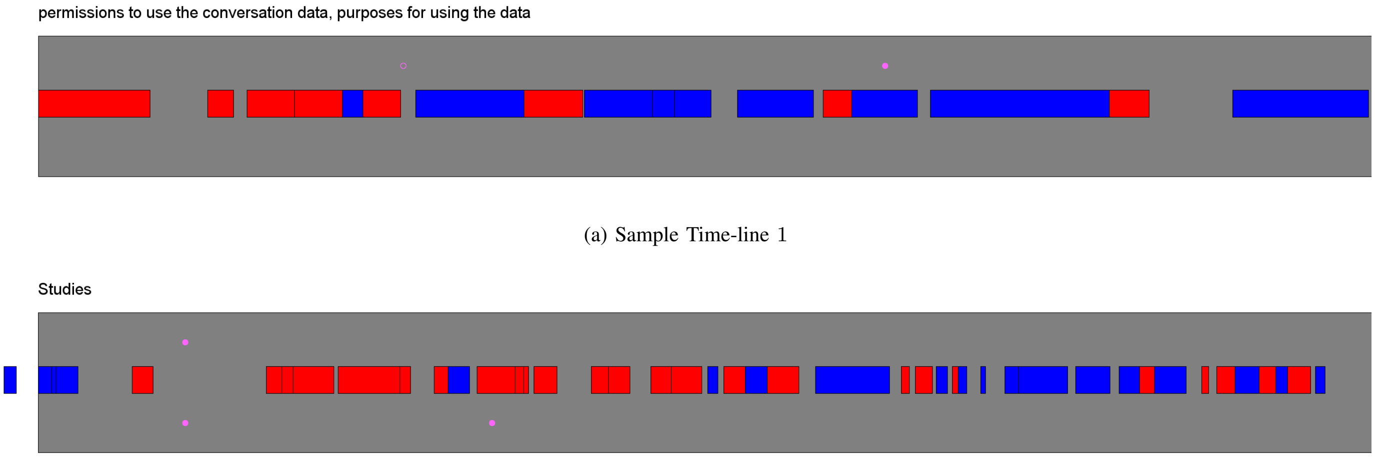
1) Word Embedding: The "word embedding" denotes a family of learning methods which does learning in form of a vector representing the word based on their co-occurrence information. Since the co-occurrence of similar words greatly is dependent on their properties, logics and semantic meanings, these vectors capture these similarities between the corresponding semantic units when they appeared in the context. Therefore, it is able to infer the distributed representation for semantic units in a vector format. Since this format is able to provide logical semantic meanings, the resulting vectors

b: breathy	56%	heavy breathing, smirk, sniff
m: mirthful	35%	fun, humorous, real laughter
e: embarrassed	4%	speaker is embarrassed, confused
p: polite	4%	polite laughter showing positive attitude towards the other speaker
o: other	1%	laughter that doesn't fit in the previous categories; acoustically unusual laughter

TABLE I: Laughter types and frequencies from [5].

	Free laughter	Speech Laugh
Breath	○	●
Embarrassed	○	●
Mirth	○	●
Derision	○	●
Polite	○	●
Other	○	●

TABLE II: Symbols about types of laughter used in Fig. 1.



(a) Sample Time-line 1

(b) Sample Time-line 2

Fig. 1: Visualisation of the time-lines of laughter and conversation

preserve the logical relation, such as "king − man + woman = queen". Since the semantic representation has been implicated in the value of the vectors, the pre-trained models using word embedding has been introduced in the applications natural language processing, language translation [20], language grounding [21], etc.

In practise, two training methods called Word2 [22] and GloVe [23] can achieve the word embedding objective, although their cost functions are different [24]. In terms of modelling, the difference between these two methods lies in different learning methods: GloVe is a "count-based" statistical model while Word2vec is a neural learning model.

In our work, we apply the Word2Vec method to learn the word embedding vector. Word2vec works as a predictive model to learn their co-occurrence vectors by a 3-layer feed-forward model[1]. In this model, the input vector and the output vector are the target word i and the context words j, respectively. The model learn the correlation between the target word and the context words in a simple way: the model tries to capture the meaningful semantic regularities by this feed-forward training. And the hidden layer, as a "by-product" after training, represents such regularities between pairs of words.

2) *LSTM Recurrent Network:* A recurrent network (RNN) is basically a feed-forward neural network with directed connecting weights forming a directed connection between the neural units in the time-domain. In such a way, the activations of the connecting neural units is not only dependent on the current inputs but also on the neural activities at the previous time-step(s). In our case, when the neural output attempts

[1] The original report of Word2vec claimed that it is a 2-layer network without the input words. Here we regard it as a 3-layer network following the convention of feed-forward network including input vector itself.

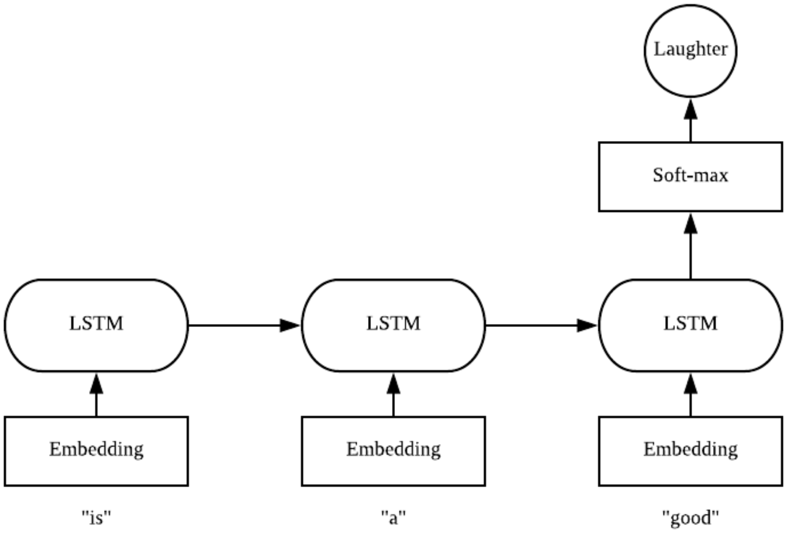


Fig. 2: The learning architecture

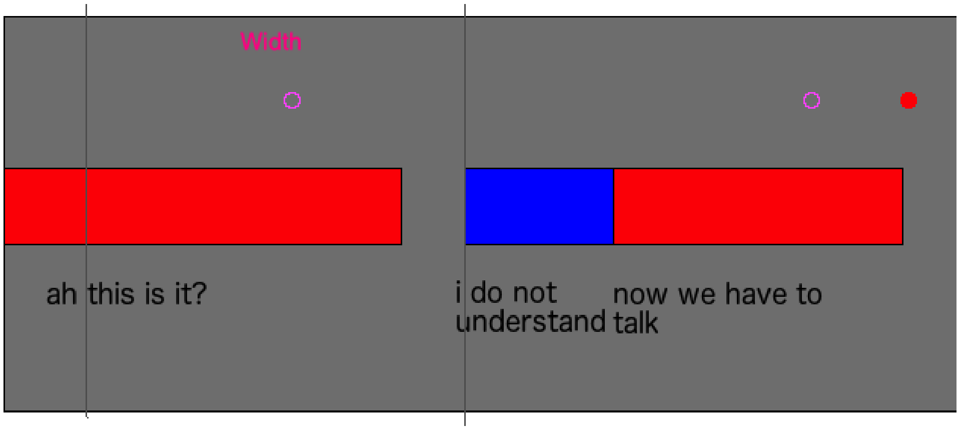


Fig. 3: Input and output of the model
Sentences below the bar are the dialogues at this time-line. The words in the dialogues (i.e. "this is it?") around the sliding window with width w are used as inputs with the format of word embedding.

to extract the temporal sequences of words, given a word embedding sequence $x = (x_1, x_2, \Delta\Delta\Delta, x_t)$ in a sentence of the dialogue, the output of the RNN $y = (y_1, y_2, \cdots, y_t)$ is computed as

$$p(y_1, y_1, \cdots, y_t) = p(y_1) \cdot p(y_2|x_1) \cdot p(y_3|x_1, x_2) \cdots p(y_t|x_1, x_2, \cdots, x_{t-1}) \quad (1)$$

In the case of RNN, the last term $p(y_t|x_1, x_2, \cdots, x_{t-1})$ can be presented as the activation of hidden unit at time t:

$$p(y_t|x_1, x_2, \cdots, x_{t-1}) = g(h(t)) \quad (2)$$

where activation function $h(t)$ is derived as Eq. 3.

$$h_t = \begin{cases} 0, t = 0 & \text{(3a)} \\ \Phi(h_{t-1}, x_t), t > 0 & \text{(3b)} \end{cases}$$

where $\Phi(\cdot)$ is a non-linear function of the recurrent units.

In our case, the long short-term memory (LSTM) units [25], [26] are utilised as the function Φ to learn the long-term dependency of the semantic values. Compared to the Vanilla RNN, the LSTM consists of various gating functions that controlled by simple element-wise operations: a LSTM has three gates: a forget gate f, an input gate i and an update gate o.

As the names imply, the forget gate determines how much of the input will be combined into the information flow with

with the previous status and the other part will be removed (or "forget"), the input gate controls how much of the input will be preserved inside the unit and the output gate defines how much of the previous memory to be output. The basic idea of using such a gating mechanism to learn long-term dependencies and eliminate the vanish gradient effects (cf. [27]). Since it was introduced, it has achieved satisfaction results in the cases which need to memorize long-term dependencies such as dialogue system [28], sentiment analysis [29] and machine translation [30].

The activation functions of each gate are given in Eq. 4:

$$f_t = \sigma(W_{fx}x_t + W_{fh}h_{t-1} + b_f) \tag{4a}$$

$$i_t = \sigma(W_{ix}x_t + W_{ih}h_{t-1} + b_i) \tag{4b}$$

$$o_t = \sigma(W_{ox}x_t + W_{oh}h_{t-1} + b_o) \tag{4c}$$

$$c_t = f_t \circ c_{t-1} + i_t \circ \sigma_c(W_{cx}x_t + W_{ch}h_{t-1} + b_c) \tag{4d}$$

$$h_t = o_t \circ \sigma_h(c_t) \tag{4e}$$

where the operator $\circ$ denotes the element-wise product. And the W and b are the corresponding weights and biases.

3) Others: On top of the output of the LSTM units y, the soft-max outputs are used for probability ($P(l)$) of the classification 12 types of laughter:

$$P(l)_j = \frac{e^{y_j}}{\sum_{j=1}^{12} e^{y_j}} \tag{5}$$

C. Experiment

Based on the models we introduced, we designed a 2-layer LSTM with embedding units to learn the co-occurence relation. The other parameters of the learning architecture are shown in the Tab. III. The Adam optimizer [31] is used for training.

Parameters	Value
Learning rate	$1e-3$
Number of hidden neurons of LSTM	50
Number of embedding units	300

TABLE III: parameters

The loss function of the raw output is defined as Eq. 6.

$$\mathcal{L} = \sum_{j=1}^{1} 2y_{j,c}log(p_j) \tag{6}$$

where the y indicate whether the class label c is the correct classification for laughter type j, p is the predicted probability observation j is of class c. We use a pre-train word2vec parameters obtained from the Google News dataset .

III. Results and discussion

A. Results

Since our assumption is that there exists a co-occurrence relation between the semantic units and the types of laughter, we change the width w of the sliding window while selecting the inputs of the conversation sentences in different training. The below figures show the learning curves of the training for the training and test sets, by changing the width w from 0.6 to 1.6 seconds, increased by 0.2 second. We can observe that the learning does not really converges, although the learning curve of the training set converges well. It indicates that there seems exist a co-occurrence relation between the semantic value of the topics in the conversation and the types of laughter. But it seems that such a relation is not strong enough to interpret the happening of different types of laughter, since the loss of the training usually fluctuates and does not converge as well as the training set.

While we change the width w of the sliding window, we can observe that the training converges best when the $w = 1s$. It may indicate that in the case of our dataset, the semantic topics within 1 second of the laughter have most co-occurrence influence to the types of it.

B. Discussion

In our intuition, a few factors can contribute to the types of laughter while people are having conversational interaction. Besides of the semantics of the topics that we investigated in our work, such as the stopping time of the conversation, the body language and even the personalities of the people who are having an interaction.

Our work based on the word embedding and recurrent neural networks has shown that there exists a relation between the semantic topics and different types of laughter, although such a relation does not seem strong in the data-set. There may be several reasons for this, and some can be further explored with future experiments. Firstly, the translated version of the dataset from Estonian to English is used for training. Therefore, the complexity and the length of different languages may have effect on the selection of the window width w. Also, only the pre-trained word-embedding layer based on Google News is used. The mentioned factors suggest that the results may differ if we utilise other options of the data-set and other pre-trained models. Nevertheless, we can still conclude that there exits certain relation between semantics of conversation and the types of the laughter, since the utterance content has the same semantics both in the original Estonian and in the translated English version of the utterance (the English translation was truthfully made to represent what the speaker said in Estonian, following the same segmentation), and the laughter occurrences are at the same time points in Estonian and English versions. In the future, however, the Estonian word2vec models can be used to verify the relation (see https://github.com/estnltk/word2vec-models).

For the development of the future work, when looking at the time-line of the laughter in the conversation, some laughter may occur randomly without much effect of the topics, especially in the case of mirthful free laugh (see Fig. 1b. Therefore, such laughter occurrences may be modelled as a point process in the conversational interaction. For instance, the Hawkes process has been used to model the interaction of social network (e.g. [32]). The properties of such interactions, such as reciprocating and sparsity, can all so fit in the laughter during conversational interaction.

Finally, another intriguing question to be answered is how the different modalities co-relate to each other when co-occurring in conversational interactions. We believe the first-encounter dataset can answer part of this question. Further investigation can also be conducted based on this dataset to find out other co-occurrence relation of other modalities.

C. Summary and Future Work

The paper has studied how dialogue topics can be learnt with neural models and confirmed that there is a correlation between topics and laughing, although the correlation is not very strong. Concerning the contributions listed in the beginning of the paper, we can conclude that the paper provides evidence for the learning of a neural model (LSTM RNN) for dialogue topics, and for the correlation of topics and laughing. Moreover, although more data is needed to confirm the model, the experiments show applicability of a small data set which is a challenge for a machine learning in general, see discussion in Trong et al. (2018) who discuss robust feature learning for dialect recognition for under-resourced languages, and propose training of an attentive network to leverage unlabeled data in a semi-supervised scenario. Another alternative, as mentioned above, is the Hawkes process, a self-exciting point process.

Topic and laughter modelling are important to understand the process of topic selection and topic switching in real-life dialogues and in particular in social robotics where the robots are expected to carry more natural interactions with the human users. We continue to explore various representations to integrate important information into our intelligent interactive robot agent, to improve naturalness of dialogue systems with the help of a laughter model, with the ultimate goal of a real-world dialogue system.

Deeper analysis will be carried out to achieve a rich representation for multimodal dialogues. We expect to integrate linguistic, acoustic and visual features into an interactive system, and explore different modalities for end-to-end conversational systems. For this we will study multimodal alignment, relations among the different modalities and hierarchical structured predictions.

ACKNOWLEDGMENT

The research was supported by Japan New Energy and Industrial Technology Development Organization (NEDO).

REFERENCES

[1] K. Jokinen, "Dialogue models for socially intelligent robots," in *International Conference on Social Robotics*. Springer, 2018, pp. 127–138.

[2] K. Jokinen and G. Wilcock, "Multimodal open-domain conversations with the nao robot," in *Natural Interaction with Robots, Knowbots and Smartphones*. Springer, 2014, pp. 213–224.

[3] K. Jokinen, S. Nishimura, K. Watanabe, and T. Nishimura, "Human-robot dialogues for explaining activities," in *Proceedings of the 9th International Workshop on Spoken Dialogue Systems Technology, Singapore*, 2018, pp. 14–16.

[4] K. Jokinen and M. McTear, "Spoken dialogue systems," *Synthesis Lectures on Human Language Technologies*, vol. 2, no. 1, pp. 1–151, 2009.

[5] K. Jokinen and T. N. Trong, "Laughter and body movements as communicative actions in interactions," *LREC 2018 Workshop on Annotation, Recognition and Evaluation of Actions*, 2018.

[6] K. Jokinen, T. N. Trong, and G. Wilcock, "Body movements and laughter recognition: experiments in first encounter dialogues," in *Proceedings of the Workshop on Multimodal Analyses enabling Artificial Agents in Human-Machine Interaction*. ACM, 2016, pp. 20–24.

[7] F. Bonin, N. Campbell, and C. Vogel, "Time for laughter," *Knowledge-Based Systems*, vol. 71, pp. 15–24, 2014.

[8] F. Bonin, "Content and context in conversations: The role of social and situational signals in conversation structure," Ph.D. dissertation, Trinity College Dublin, 2016.

[9] D. Alvarez-Melis and M. Saveski, "Topic modeling in twitter: Aggregating tweets by conversations." *ICWSM*, vol. 2016, pp. 519–522, 2016.

[10] D. M. Blei, "Probabilistic topic models," *Communications of the ACM*, vol. 55, no. 4, pp. 77–84, 2012.

[11] C. C. Aggarwal and C. Zhai, "A survey of text clustering algorithms," in *Mining text data*. Springer, 2012, pp. 77–128.

[12] W. L. Chafe, *The importance of not being earnest: The feeling behind laughter and humor*. John Benjamins Publishing, 2007, vol. 3.

[13] E. Holt, "Conversation analysis and laughter," *The encyclopedia of applied linguistics*, 2012.

[14] E. Goffman, *Frame analysis: An essay on the organization of experience.* Harvard University Press, 1974.

[15] J.-A. Bachorowski, M. J. Smoski, and M. J. Owren, "The acoustic features of human laughter," *The Journal of the Acoustical Society of America*, vol. 110, no. 3, pp. 1581–1597, 2001.

[16] J. Trouvain, "Segmenting phonetic units in laughter," in *Proc. 15th International Conference of the Phonetic Sciences, Barcelona, Spain*, 2003, pp. 2793–2796.

[17] H. Tanaka and N. Campbell, "Acoustic features of four types of laughter in natural conversational speech," in *Proc. 17th International Congress of Phonetic Sciences (ICPhS), Hong Kong*, 2011, pp. 1958–1961.

[18] K. Jokinen and S. Tenjes, "Investigating engagement intercultural and technological aspects of the collection, analysis, and use of estonian multiparty conversational video data," in *Proceedings of the Language Resources and Evaluation Conference (LREC-2012*. Citeseer, 2012.

[19] K. P. Truong and D. A. Van Leeuwen, "Automatic discrimination between laughter and speech," *Speech Communication*, vol. 49, no. 2, pp. 144–158, 2007.

[20] Y. Qi, D. S. Sachan, M. Felix, S. J. Padmanabhan, and G. Neubig, "When and why are pre-trained word embeddings useful for neural machine translation?" *arXiv preprint arXiv:1804.06323*, 2018.

[21] J. Zhong, T. Ogata, A. Cangelosi, and C. Yang, "Understanding natural language sentences with word embedding and multi-modal interaction."

[22] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," *arXiv preprint arXiv:1301.3781*, 2013.

[23] J. Pennington, R. Socher, and C. D. Manning, "Glove: Global vectors for word representation." in *EMNLP*, vol. 14, 2014, pp. 1532–1543.

[24] T. Shi and Z. Liu, "Linking glove with word2vec," *arXiv preprint arXiv:1411.5595*, 2014.

[25] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[26] F. A. Gers, J. Schmidhuber, and F. Cummins, "Learning to forget: Continual prediction with lstm," 1999.

[27] J. Zhong, A. Cangelosi, and T. Ogata, "Toward abstraction from multi-modal data: empirical studies on multiple time-scale recurrent models," in *Neural Networks (IJCNN), 2017 International Joint Conference on*. IEEE, 2017, pp. 3625–3632.

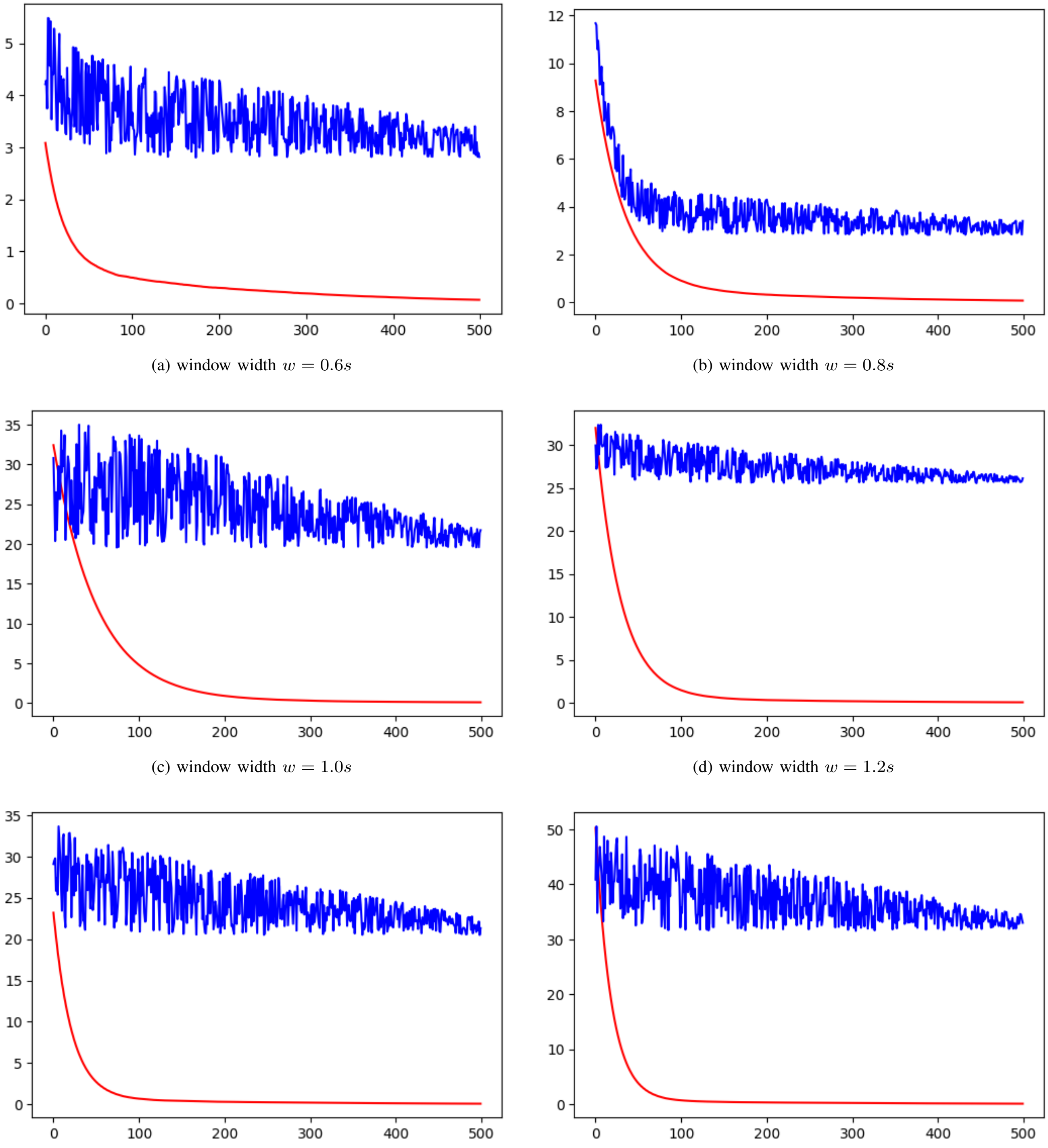
[28] O. Vinyals and Q. Le, "A neural conversational model," *arXiv preprint arXiv:1506.05869*, 2015.

[29] A. M. Dai and Q. V. Le, "Semi-supervised sequence learning," in *Advances in Neural Information Processing Systems*, 2015, pp. 3079–3087.

[30] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," in *Advances in neural information processing systems*, 2014, pp. 3104–3112.

[31] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[32] X. Miscouridou, F. Caron, and Y. W. Teh, "Modelling sparsity, heterogeneity, reciprocity and community structure in temporal interaction data," in *Advances in Neural Information Processing Systems*, 2018, pp. 2345–2354.



(a) window width $w = 0.6s$

(b) window width $w = 0.8s$

(c) window width $w = 1.0s$

(d) window width $w = 1.2s$

(e) window width $w = 1.4s$

(f) window width $w = 1.6s$

Fig. 4: Learning curve with different window width w.
The red curve represents the loss of the training set, and the blue curves represents the loss of the evaluation set.