ConvLogMiner: A Real-Time Conversational Lifelog Miner

Pei-Wei Kao[1] , **An-Zi Yen**[1] , **Hen-Hsen Huang**[2,3] and **Hsin-Hsi Chen**[1,3]
[1]Department of Computer Science and Information Engineering,
National Taiwan University, Taipei, Taiwan
[2]Institute of Information Science, Academia Sinica
[3]MOST Joint Research Center for AI Technology and All Vista Healthcare, Taipei, Taiwan
{pwgao, azyen, hhhuang}@nlg.csie.ntu.edu.tw, hhchen@ntu.edu.tw

Abstract

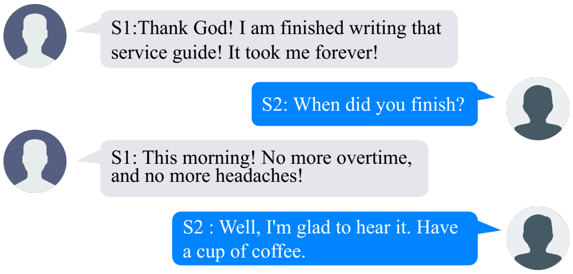
This paper presents a conversational lifelog mining system, ConvLogMiner, which detects personal life events from the human online conversation in real-time. Given a daily conversation of two speakers, ConvLogMiner identifies the new life events specific to each speaker that occur in the latest utterances. The lifelogs mined by our system are useful to provide complementary information to support lifestyle analysis and memory assistance service.

1 Introduction

Lifelogging, a process of actively capturing the daily life experiences of an individual and storing them digitally, has gained a lot of attention in recent years [Yen *et al.*, 2021a]. With the advance of digital technology, people are used to sharing their life by taking photos/videos or writing social media posts. The collection of lifelogs can be used for analyzing our lifestyle [Doherty *et al.*, 2011; Maekawa, 2013] and supporting memory assistance [Rahman *et al.*, 2018; Caros *et al.*, 2020; Yen *et al.*, 2021b]. In addition, backtracking where we have been to and who we have been in contact with enables identifying the probability that we exposure to a source of infection [Bengio *et al.*, 2020].

Previous researches focus on extracting life events from social media posts [Li *et al.*, 2014; Yen *et al.*, 2019a; Yen *et al.*, 2019b], videos and pictures [Song *et al.*, 2014; Yoshikawa *et al.*, 2018], but rarely discuss on conversation. In our daily life, we not only exchange information through communication, but also share our past experiences with others. That is to say, conversations provide complementary information about an individual's life events that can be used for supporting a variety of lifelogging applications.

Some works have been done for identifying the speaker's attributes such as gender, age, or relationship status from conversation transcripts [Garera and Yarowsky, 2009; Welch *et al.*, 2019a; Welch *et al.*, 2019b]. Tigunova et al. [2021] propose a web demonstration platform based on a personal knowledge extraction model, CHARM [Tigunova *et al.*, 2020], to infer personal attributes like hobby and profession by asking several questions with the user. In contrast to designing a chatbot to interact with users for collecting personal attributes, we aim to detect daily life events (e.g., where did the user go or what they ate) of each speaker in a human-human conversation.



S1:Thank God! I am finished writing that service guide! It took me forever!

S2: When did you finish?

S1: This morning! No more overtime, and no more headaches!

S2 : Well, I'm glad to hear it. Have a cup of coffee.

Figure 1: Example of Human Conversation.

Figure 1 shows an example of people talking about their life experiences in a conversation. The life events of speaker 1 (S1) are that S1 finished the service guide and drank a cup of coffee. And the life event of S2 is that she/he heard S1 had finished the work. Two main challenges to be tackled are shown as follows.

1. In natural conversation, people often use pronouns to describe objects that were previously mentioned.
2. Inferencing speaker's life event may require information from other speaker's utterances.

For example, understanding that "this morning" in the third utterance represents the time of S1 finishing the work depends on the question asked by S2. On the other hand, the implicit life event of S1 drinking a cup of coffee has to be inferred from the S2's last utterance. In other words, handling ellipsis and co-reference in conversations is an important issue to be dealt with. In addition, being aware of the speakers and comprehending the entire conversation is a crucial problem.

In this work, we construct a pilot lifelog miner system, ConvLogMiner[1], demonstrating how to detect life events from human conversation in real-time. ConvLogMiner is capable of detecting personal life events from the online human-human conversation. To address the aforementioned challenging issues, we propose three modules which encode the contextual information of the conversation under different strategies. The life events of speakers detected by our system are based on the predictions of the three modules. The

[1] https://nlg.tw/ConvLogMiner

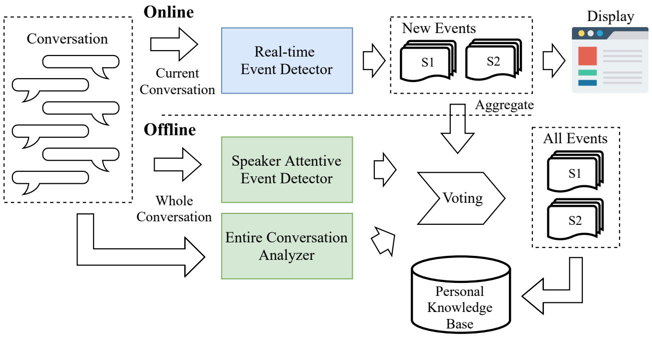


Figure 2: System Overview of ConvLogMiner.

details will be described in Section 2. The contributions of this work are threefold as follows.

1. This work introduces a new direction of lifelogging to facilitate future research.
2. We construct a novel system to demonstrate life event detection in the conversation in real-time. Experimental results show promising performances are achieved.
3. In addition to the demonstration system, we also construct a human-annotated conversational lifelog dataset.

2 ConvLogMiner

An overview of our system is illustrated in Figure 2. ConvLogMiner provides two kinds of services. One is the online real-time detection service, which provides instant feedback of the detected new life events in the current turn. The other one is the offline life event analysis service, which refines the detection results and stores them into the personal knowledge bases of the corresponding speakers for further lifelogging applications. Our system is composed of three modules that follow the framework of XLNet [Yang *et al.*, 2019] with different encoding strategies. For online detection, the core module is the real-time event detector denoted in blue that focuses on mining lifelogs of each speaker in the last utterance. As for offline analysis, we construct two auxiliary modules denoted in green to refine the results of the online detection. One auxiliary module is the speaker attentive event detector which aims at detecting all the life events of a specific speaker according to her/his utterances only. The other auxiliary module is the entire conversation analyzer which recognizes all the life events of a speaker based on all utterances in a conversation. The detail of three modules are described as follows.

Framework of XLNet. Given a speaker S and a conversation $C = \{t_1, t_2, ..., t_n\}$, where t_i denotes the i-th turn in C and n is the total number of turns. The i-th turn $t_i = \{s_j, w_{i1}, w_{i2}, ..., w_{im}\}$ is an uninterrupted stream of speech from speaker s_j, where $j \in \{1, 2\}$ represents S1 or S2, and m is the number of tokens in t_i. We concatenate S and C with special tokens `[CLS]` and `[SEP]` as the input sequence. After fed into XLNet, we take the final hidden vector of `[CLS]` as the conversation embedding $\mathbf{h}_C$.

Our goal is to detect the life event types, which is viewed as a multi-label classification task. We use a sigmoid layer $\sigma(W^C\mathbf{h}_C)$ to predict the event types y_C of C by computing the probability of $P(y_C|S_j, C)$, where $y_C \in \{0, 1\}^\alpha$ (α is the set of the event types), and $W^C \in \mathbb{R}^H$ (H is the hidden size). We utilize binary cross-entropy as loss function. On testing stage, if the probability of the event type et in α is higher than threshold 0.5, we consider that et occurs in C.

Real-Time Event Detector. To achieve online detection of life events, we sequentially perform the real-time event detector for each turn. Since there are two speakers in a conversation, we perform $2\times n$ times. We encode k turns in backward ordering as input conversation $C_k = \{t_k, t_{k-1}, ..., t_1\}$, where t_k represents the last turn. Here, we set the maximum turn window size as 6. Follow the framework of XLNet, this module is aimed at detecting the new personal life events triggered by the last turn in the conversation.

Speaker Attentive Event Detector. For excluding the noise that may be introduced from the other speaker's utterances, this module focuses on mining all life events based on the utterances of a specific speaker. We take the turns of the given speaker S as input conversation $C_j = \{t_{j_1}, t_{j_2}, ..., t_{j_q}\}$, where t_{j_i} denotes the i-th turn of the given speaker S_j, and q is total number of turns of S_j.

Entire Conversation Analyzer. Considering some life events still need to be inferred by other speaker's utterances, we also take the entire conversation C as input to detect all mentioned life events of each speaker.

To be specific, in online service, ConvLogMiner will iteratively detect whether new life events occur in the last turn of the current conversation by the real-time event detector showcasing the initial detection results. When the conversation is over, we provide the offline service that allows the user to check all life events detected by ConvLogMiner, which is performed and refined by three modules with the ensemble mechanism. For the refinement, we first aggregate the results of the real-time event detector to obtain all predicted event types in the whole conversation, since the real-time event detector predicts each speaker's life events in each turn incrementally. Then, we convert the entire conversation into the input sequences of the two auxiliary modules, and feed them into the modules to obtain the predictions. Finally, the system will ensemble the results of the three modules by majority voting as all life events of each speaker, which will be stored in her/his personal knowledge base.

3 Dataset Construction

Since there are no available conversational datasets for lifelog mining, we extend a multi-turn dialog dataset, DailyDialog [Li *et al.*, 2017], with life event annotation to construct a dataset for life event detection from daily dialog, which is referred as DiaLog.[2] We define the personal life event as the activity of daily living that relates to a specific individual.

We collect 600 conversations with four to six utterances from four different topics, including Ordinary Life, Attitude & Emotion, Relationship and Tourism. For the annotation, we invite three annotators who major in linguistics. For each

[2] http://nlg.csie.ntu.edu.tw/nlpresource/DiaLog/

Event Types	Related Predicates	Frequency
Perception_experience	watch, see, hear	17.62 %
Activity	go, play	10.47 %
Request	order, request	10.22 %
Contacting	call	9.32 %
Transporting	fly, drive, take	8.56 %

Table 1: Most frequent event types and their related predicates.

turn of a conversation, an annotator is asked to label life events appearing in the conversation by the following steps:

1. Whether the last utterance triggers one or more life events in the conversation.
2. Specify the actor and the predicate of each life event.
3. According to FrameNet ontology [Fillmore *et al.*, 2002], select a suitable frame name as event type for each predicate.

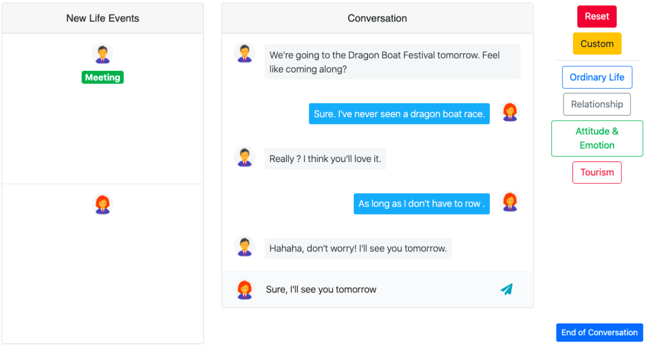
To verify the quality of annotations, we select 40 conversations labeled by all annotators to measure agreement. The inter annotator agreement, krippendorff's alpha, is 0.81, which achieves a reliable score. Finally, 21 unique event types are chosen and the total number of annotated events is 780. The percentage of conversations with and without life events are 71.83% and 28.17%, respectively. Table 1 shows the top 5 frequent life event types in DiaLog. We find that speakers tend to share what they saw, what they heard, and where they have been to in daily life. Besides, speakers often express their needs when communicating with others.

4 Demonstration

Figure 3 shows the screenshot of ConvLogMiner's online service page. We present the conversation scenario as a chat platform. In the block of Conversation, users can input messages on behalf of different speakers to imitate a daily talk with others. After sending a message, ConvLogMiner will predict life events of both speakers triggered by the new message with the real-time event detector and show the results in the block of New Life Events. Note that if users enter a message on behalf of the same speaker in the last line, our system will consider them as the same turn of the speaker. While pressing the blue button "End of Conversation" on the right bottom corner, the system will switch to offline mode and automatically start the refinement. The offline analysis page will show the entire conversation, the predictions of the three modules, and the results of the ensemble mechanism for storing in the personal knowledge bases for the two speakers.

5 Evaluation

The training and test sets contain 500 and 100 instances, respectively. We adopt the pre-trained model, *xlnet-base-cased*, and set the learning rate to 3e-5 and the maximum input length to 512 tokens to train our models.



We're going to the Dragon Boat Festival tomorrow. Feel like coming along?

Sure. I've never seen a dragon boat race.

Really ? I think you'll love it.

As long as I don't have to row .

Hahaha, don't worry! I'll see you tomorrow.

Sure, I'll see you tomorrow

Figure 3: Screenshot of ConvLogMiner's Online Service Page.

Model	F1(%)	P(%)	R(%)
XLNet$_{\text{Aggre}}$	44.21	43.31	45.15
XLNet$_{\text{SelfTurns}}$	42.33	43.04	41.64
XLNet$_{\text{AllTurns}}$	39.56	33.52	**48.27**
Ensemble	**45.30**	**43.92**	46.78

Table 2: Experimental results.

We evaluate the experimental results by macro F-score (F1), precision (P), and recall (R). Table 2 shows the results of models used in the three modules and the refinement by using the ensemble mechanism. **XLNet$_{\textbf{Aggre}}$** denotes the aggregation of real time event detector. **XLNet$_{\textbf{SelfTurns}}$** represents model in the speaker attentive event detector. **XLNet$_{\textbf{AllTurns}}$** is the model in the entire conversation analyzer. **Ensemble** means the result of the refinement by three modules.

In Table 2, XLNet$_{\text{Aggre}}$ achieves 44.21% macro F-score, which outperforms the other models used in two auxiliary modules. That is, detecting life events in the i-th turn sequentially is more appropriate than considering the entire conversation. However, **XLNet$_{\textbf{AllTurns}}$** achieves the highest recall of 48.27%. For taking advantage of the three models, the ensemble mechanism is conducted. Experimental results show that utilizing the ensemble mechanism is effective in obtaining a better refinement result from the three models, and achieves the highest F1 score of 45.30%.

6 Conclusion and Future Work

This work demonstrates a pilot system ConvLogMiner to mine lifelogs of speakers in a conversation. Compared with previous work of identifying personal attributes only from the conversation, our system detecting the life events of each speaker is expected to be useful to further lifelogging applications. Note that we only focus on detecting event types at the current stage. We plan to extract the subject, predicate, and object of a life event in a triple form to store in the personal knowledge base for easy organization. Considering that the system detecting the life events of users from the conversation will potentially endanger their privacy, the investigation of a privacy-aware lifelogging framework becomes an emerging research issue that we also leave as future work.

Acknowledgments

This research was partially supported by Ministry of Science and Technology, Taiwan, under grants MOST-109-2634-F-002-034- and MOST-110-2634-F-002-028-.

References

[Bengio *et al.*, 2020] Yoshua Bengio, R. Janda, Y. W. Yu, Daphne Ippolito, Max Jarvie, D. Pilat, Brooke Struck, Sekoul Krastev, and A. Sharma. The need for privacy with public digital contact tracing during the covid-19 pandemic. *The Lancet. Digital Health*, 2:e342 – e344, 2020.

[Caros *et al.*, 2020] Mariona Caros, Maite Garolera, Petia Radeva, and Xavier Giro-i Nieto. Automatic reminiscence therapy for dementia. In *Proceedings of the 2020 International Conference on Multimedia Retrieval*, pages 383–387, 2020.

[Doherty *et al.*, 2011] Aiden R Doherty, Niamh Caprani, Vaiva Kalnikaite, Cathal Gurrin, Alan F Smeaton, Noel E O'Connor, et al. Passively recognising human activities through lifelogging. *Computers in Human Behavior*, 27(5):1948–1958, 2011.

[Fillmore *et al.*, 2002] Charles J Fillmore, Collin F Baker, and Hiroaki Sato. The framenet database and software tools. In *Proceedings of the Third International Conference on Language Resources and Evaluation (LREC'02)*, pages 1157–1160, 2002.

[Garera and Yarowsky, 2009] Nikesh Garera and David Yarowsky. Modeling latent biographic attributes in conversational genres. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*, pages 710–718, 2009.

[Li *et al.*, 2014] Jiwei Li, Alan Ritter, Claire Cardie, and Eduard Hovy. Major life event extraction from twitter based on congratulations/condolences speech acts. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1997–2007, 2014.

[Li *et al.*, 2017] Yanran Li, Hui Su, Xiaoyu Shen, Wenjie Li, Ziqiang Cao, and Shuzi Niu. DailyDialog: A manually labelled multi-turn dialogue dataset. In *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 986–995, Taipei, Taiwan, November 2017. Asian Federation of Natural Language Processing.

[Maekawa, 2013] Takuya Maekawa. A sensor device for automatic food lifelogging that is embedded in home ceiling light: A preliminary investigation. In *2013 7th International Conference on Pervasive Computing Technologies for Healthcare and Workshops*, pages 405–407. IEEE, 2013.

[Rahman *et al.*, 2018] Md Abed Rahman, AM Esfar E Alam, Md Hasan Mahmud, and Md Kamrul Hasan. Towards a smartphone based lifelogging system for reminiscence. *Journal of Engineering and Technology*, 14(1), 2018.

[Song *et al.*, 2014] Sibo Song, Vijay Chandrasekhar, Ngai-Man Cheung, Sanath Narayan, Liyuan Li, and Joo-Hwee Lim. Activity recognition in egocentric life-logging videos. In *Asian Conference on Computer Vision*, pages 445–458. Springer, 2014.

[Tigunova *et al.*, 2020] Anna Tigunova, Andrew Yates, Paramita Mirza, and Gerhard Weikum. Charm: Inferring personal attributes from conversations. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 5391–5404, 2020.

[Tigunova *et al.*, 2021] Anna Tigunova, Paramita Mirza, Andrew Yates, and Gerhard Weikum. Exploring personal knowledge extraction from conversations with charm. In *Proceedings of the 14th ACM International Conference on Web Search and Data Mining*, pages 1077–1080, 2021.

[Welch *et al.*, 2019a] Charles Welch, Verónica Pérez-Rosas, Jonathan K. Kummerfeld, and Rada Mihalcea. Learning from personal longitudinal dialog data. *IEEE Intelligent systems*, 34(4), 2019.

[Welch *et al.*, 2019b] Charles Welch, Verónica Pérez-Rosas, Jonathan K. Kummerfeld, and Rada Mihalcea. Look who's talking: Inferring speaker attributes from personal longitudinal dialog. In *Proceedings of the 20th International Conference on Computational Linguistics and Intelligent Text Processing (CICLing)*, La Rochelle, France, 2019.

[Yang *et al.*, 2019] Z. Yang, Zihang Dai, Yiming Yang, J. Carbonell, R. Salakhutdinov, and Quoc V. Le. Xlnet: Generalized autoregressive pretraining for language understanding. In *NeurIPS*, 2019.

[Yen *et al.*, 2019a] An-Zi Yen, Hen-Hsen Huang, and Hsin-Hsi Chen. Multimodal joint learning for personal knowledge base construction from twitter-based lifelogs. *Information Processing & Management*, 57(6), 2019.

[Yen *et al.*, 2019b] An-Zi Yen, Hen-Hsen Huang, and Hsin-Hsi Chen. Personal knowledge base construction from text-based lifelogs. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 185–194, 2019.

[Yen *et al.*, 2021a] An-Zi Yen, Hen-Hsen Huang, and Hsin-Hsi Chen. Ten questions in lifelog mining and information recall. In *Proceedings of the 2021 International Conference on Multimedia Retrieval*, 2021.

[Yen *et al.*, 2021b] An-Zi Yen, Hen-Hsen Huang, and Hsin-Hsi Chen. Unanswerable question correction in question answering over personal knowledge base. In *Thirty-Fifth AAAI Conference on Artificial Intelligence (AAAI-21)*, 2021.

[Yoshikawa *et al.*, 2018] Yuya Yoshikawa, Jiaqing Lin, and Akikazu Takeuchi. Stair actions: A video dataset of everyday home actions. *arXiv preprint arXiv:1804.04326*, 2018.