Linguistic Features of Clients and Counselors for Early Detection of Mental Health Issues in Online Text-based Counseling

Kazuhiro Shidara[1], Hiroki Tanaka[1], Rumiko Asada[2], Kayo Higashiyama[2], Hiroyoshi Adachi[3], Daisuke Kanayama[3],
Yukako Sakagami[3], Takashi Kudo[3], Satoshi Nakamura[1]

1) Nara Institute of Science and Technology, 2) Osaka Prefectural Goverment, 3) Osaka University E-mail: shidara.kazuhiro.sc5@is.naist.jp

Japan Science and Technology Agency CREST

INTRODUCTION

■ SNS counseling [Sugihara, 2020]
- Remote counseling with messaging apps (e.g., LINE).

■ Problem
- Recognizing the types of user concerns has a heavy workload for counselors.

■ Related work
- There are some studies classify that clients are satisfied in session or not (e.g. [Altoff et al., 2016]).

->There is a lack of research to classify the type of clients' issues to assist counselors.

AIM

1. To investigate text features with classification performance and high explanatory.
2. To investigate whether counselors' or clients' data contribute to classification performance.
3. To investigate the possibility of early classification after the start of the session.

TASK

■ Text classification for the two client types
- Mental health issues (N=339)
- Other issues (N=635)

Number of clients in each trouble

Total number of issues		974
Mental health issues	Mental health	384
	Suicide thoughts	5
Other issues	Family	162
	Physical health	131
	Work	103
	Finance	81
	Sexual relationships	37
	Prejudice concerning COVID-19	5
	School	17
	Others	49

*Annotation of issues was conducted by licensed counselors affiliated with the data provider.

METHOD

■ Data setting
- Only Clients, Only Counselors, Clients and counselors

■ Feature setting
- **TF-IDF:** A method considers the importance of words in a sentence to describe its features. The frequency of a word's use in a single document and its frequency of use in all the documents are used to calculate a word that represents each document.
- **Contextualized document embeddings (Embeddings):** For context-aware document embedding, we used a large-scale language model called Bidirectional Encoder Representations from Transformers (BERT).
- **Latent Dirichlet Allocation (LDA):** A method to compute latent topics with unsupervised learning.
- **LIWC:** A linguistic resource for categorizing words from a psychological perspective.
- **Statistics:** The following descriptive statistics.

List of statistics and mean values (Cl: Client, Co: Counselor)

		Mental health issues	Other issues	p	Hedge's g (Effect size)
Tokens	Cl	950	879	0.061	0.117
	Co	959	868	**0.004**	0.191
Types	Cl	197	186	0.054	0.121
	Co	181	171	**0.008**	0.172
Types/Tokens	Cl	0.247	0.267	**<0.001**	-0.241
	Co	0.208	0.227	**<0.001**	-0.325
Types/$\sqrt{Tokens}$	Cl	6.55	6.53	0.706	0.024
	Co	5.96	6.00	0.439	-0.050

		Mental health issues	Other issues	p	Hedge's g (Effect size)
Seconds		5476	4848	**<0.001**	0.240
Number of messages	Cl	32.1	29.4	0.090	0.105
	Co	27.8	24.7	**<0.001**	0.219
Seconds /Messages	Cl	100	103	0.363	-0.057
	Co	100	104	0.217	-0.081
Ratio of the number of cl. and co. messages		1.11	1.20	0.377	-0.055

- **All features:** A combination of all features.

■ Classifier: Logistic regression

RESULTS

■ The best performance was shown in the following condition.
- Data: Clients' message
- Feature: TF-IDF

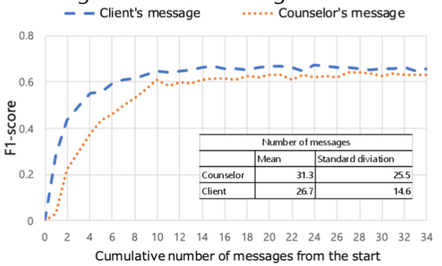
Performance of the detection models

Data used for training and test	Feature	Precision	Recall	F1-score	AUC
Client	TF-IDF	**0.614**	0.681	**0.646**	**0.703**
	Embeddings	0.597	**0.698**	0.643	0.698
	LDA	0.393	0.600	0.475	0.503
	LIWC	0.391	0.500	0.439	0.501
	Statistics	0.393	0.450	0.420	0.502
	All features	0.452	0.479	0.465	0.553
Counselor	TF-IDF	**0.556**	**0.633**	**0.592**	**0.651**
	Embeddings	0.447	0.538	0.488	0.550
	LDA	0.401	0.503	0.446	0.505
	LIWC	0.399	0.500	0.444	0.503
	Statistics	0.414	0.428	0.421	0.515
	All features	0.448	0.459	0.453	0.543
Client and Counselor	TF-IDF	**0.576**	**0.668**	**0.618**	**0.678**
	Embeddings	0.552	0.661	0.602	0.661
	LDA	0.389	0.400	0.394	0.501
	LIWC	0.388	0.408	0.398	0.501
	Statistics	0.380	0.498	0.431	0.492
	All features	0.450	0.479	0.464	0.554

We investigated the early detection performance and the weights of the features in the best performance condition (Clients' message,TF-IDF).

■ It shows almost the maximum estimated performance in the early stage.

Relationship between number of dialogue messages used for training and F1 scores



■ The model has been able to utilize words that suggest mental status, such as "Palpitations," "Scary," "Though," "Calm," and "Fear."

Top 10 words with high weighting

Client	Counselor	Client and counselor
Fear	Palpitations	Scary
Scary	Scary	Look
Palpitations	Firmly	Infection
Infection	Calm	Palpitations
Television	Out	Tough
Look	When	Calm
Out	Caution	News
Tough	Doctor	Out
Many	Important	Television
Increase	Great	Increase

CONCLUSIONS

1. TF-IDF shows the best performance and the high-weighted words were explainable.
2. The model achieved higher classification performance under conditions with clients' data.
3. The clients' data showed high classification performance relatively early stage.

REFERENCES

- Y. Sugihara and T. Miyata, The SNS Counseling Handbook. Seishin Shobo,Ltd., 2020.
- T. Althoff, K. Clark, and J. Leskovec. Large-scale analysis of counseling conversations: An application of natural language processing to mental health. Transactions of the Association for Computational Linguistics, 2016, 4: 463-476.

ACKNOWLEDGEMENTS

- This work was supported by JST CREST Grant Number JPMJCR19A5, Japan.
- We thank Osaka Prefectural Government and Kansai Counseling Center for their cooperation.