

## 来談者中心療法を用いたカウンセリングシステムの性能比較

堀井翼<sup>†1</sup> 櫻井義尚<sup>†1</sup> 櫻井恵里子<sup>†2</sup> 鶴田節夫<sup>†3</sup>

†1 明治大学大学院先端数理科学研究科

†2 西武文理大学サービス経営学部

†3 東京電機大学情報環境科学研究科

## 1. はじめに

近年、様々な苦悩を抱えている人が多い。特に日本のIT労働者の60%が精神的なストレスを抱え、十分な仕事を出来ずにいる[1]。そして、IT開発プロジェクトの成功率は30%にとどまっている[2]。一方で、カウンセラーの数が少ない。そのため、人々は自身の問題や苦しみを拭いきれずにいる。故に、人間のカウンセラーに代替するシステムの開発が必要である。対話システムとしては、1966年にELIZAというシステムが提案された[3]。しかしELIZAは実社会で来談者の問題解決をするには至らなかった。そのため、我々は2013年に、CRECA(Context Respectful Counseling Agent)と呼ばれるELIZAのようなミラーリングを主機能としつつ、実社会での問題解決を目指した対話システムを提案した[4]。しかしながら、CRECAには2点の課題が指摘された。1つ目は、それぞれの問題に応じて、様々な分野に対する知識構築が必要であり、汎用性が低いこと。2つ目は、インターフェースがテキストであること。そこで、我々は音声対話によるVCAを提案する。本稿ではELIZA, CRECA, VCAの3つの性能を比較する実験を行い、考察する。

## 2. 関連研究

関連研究は、カウンセリングと対話システムの2分野に分けられる。

## (1) カウンセリング

来談者中心療法がロジャースによって提案されている[10]。カウンセラーが来談者の意見を肯定し、受け止めることで、来談者の自己解決を促す療法だ。私たちのエージェントではこの手法を用いる。

## (2) 対話システム

ELIZAが来談者中心療法の第一ワークである[3]。ELIZAは来談者の発言をミラーリングすることで、会話を促進した。しかしELIZAが実社会で、人の問題を解決するには機能が不十分であった。そのため2013年にCRECAが提案された[4]。CRECAでは、カウンセリングプロセスを問題発見フェイズと問題解決フェイズ2つに分けた。問題発見フ

ェイズではテンプレート式でカウンセラーと来談者の自己紹介を行い、それから問題の種類を特定する。問題解決フェイズでは、ミラーリング機能に加え、感情認識機能やContext Based Reasoning (CBR) [5]に基づいて、現在の対話が問題解決とそれないようにする機能を実装した。しかしCRECAの問題解決にはそれぞれの問題に応じて、様々な分野に対する知識構築が必要であった。そのため汎用的に問題解決できない。

## 3. システム設計

VCAのシステムでは、CRECAから2点の改善を行った。1つ目は、対話プロセスの汎用化。2つ目はインターフェースの改善である。

## 3.1 対話プロセス

VCAの根本的な対話プロセスはCRECAと同じである。来談者が発言すると、それをmecabで形態素解析する。それから感情辞書に基づいて10種の感情を認識する[6]。そしてCBRを用いながら、応答パターンとマッチングをとり、返答をする。上記のCRECAの設計から、VCAでは問題発見フェイズを無くし、かつ、知識構築を行わない設計にすることで、より自然な会話の中でのカウンセリングを実現し、より様々な分野の問題解決に対応できるようにした。

## 3.2 インターフェース設計

VCAのインターフェースは、2Dアバターと音声対話を用いている。図1にそのアバターを示す。音声対話はGoogle Chrome上のアバターを通して、Google Cloud speech APIで音声を文字化。その後CRECAのプロセスと同様に応答文を作成し(3.1参照)、Google Speech Synthesizerを用いて音声出力している。さらに、アバターは“うんうん”と、相づちを打ちながら、頷き動作を行なう。理由は、カウンセリングシステムが来談者に対して頷きと相づちを行なうことに対する効果が証明されているからだ[7]。WebサーバーはPythonを用いてNginx上で起動している。

†1 Tsubasa Horii, Yoshitaka Sakurai, Graduate School of Advanced Mathematical Sciences, Meiji University  
†2 Eriko Sakurai, Faculty of Service Management, Bunri University of Hospitality

†3 Setsuo Tsuruta, School of Information Environment, Tokyo Denki University



図 1

#### 4. 実験・評価

VCA の評価は ELIZA、CRECA と比較することで行なう。3 つの順番はランダムに行なった。評価方法は 3 つ。アンケートと対話時間、それから対話数である。10 人の学生を実験参加者とした。アンケートは過去に提案されているものをを用いた[8]。22 項目の 7 段階リッカート尺度[9]からなり、22 項目中、10 項目はカウンセラーへの信頼や親しみに関する質問(TRUST)、12 項目は問題の自己解決に関する質問(SELF-AWARENESS)で構成されている。7 段階のうち、7 が高評価である。

実験結果を表 1 に示す。アンケート結果からフリードマン検定[10, 11, 12]を有意水準 1%で行ない、ウィルコクソンの符号順位検定[13]を有意水準 5%で行った。フリードマン検定では有意差を確認できた。ウィルコクソンの符号順位検定では、TRUST、SELF-AWARENESS 共に、ELIZA-CRECA 間と ELIZA-VCA 間では有意差を確認できた。しかし、CRECA-VCA 間では確認できなかった。

対話時間と対話数を考慮すると、VCA は CRECA より短時間かつ、少ない対話で CRECA と同様以上の評価を得ていることが分かる。

表 1

	Questionnaire		Time-Average (seconds)	The number of dialog interactions-Average
	TRUST-Average (1-7)	SELF AWARENESS-Average (1-7)		
ELIZA	<b>2.07</b> <0.99, 0.98>	<b>2.81</b> <1.09, 1.15>	<b>429.4</b> <201.7, 202.6>	<b>16.8</b> <5.86, 6.18>
CRECA	<b>3.78</b> <1.40, 1.39>	<b>4.46</b> <1.11, 1.17>	<b>849.5</b> <485.45, 511.6> (605.0, 244.5)	<b>26.1</b> <10, 10.57> (20.3, 5.8)
VCA	<b>4.07</b> <1.20, 1.20>	<b>4.46</b> <1.05, 1.11>	<b>379.8</b> <232.76, 245.34>	<b>14.6</b> <5.97, 6.31>

#### 5. 結論

本稿では VCA を構築し、既存の対話システムとの性能比較実験を行なった。その結果、VCA が高い評価を得た。VCA では音声認識を採用しているため、聞き取りミスが確認されたが、それでもなお、VCA は CRECA と同等以上の評価を得た。これは CRECA から VCA への対話フロー改善

とインターフェース改善が良い結果を導いたことを示している。さらに VCA は事前知識の構築が必要なく、実社会に近いより自然な会話の中での問題解決を実現する。これは高齢者をはじめとした、デジタルデバイドへのアプローチの有用性も示している。VCA と CRECA の有意差を確認できなかった理由として、CRECA の問題発見フェイズが挙げられる。問題発見フェイズは枠にはまった会話の中で自己紹介や問題特定が行なわれるため、実験参加者からの評判が良かった。以上のことから、今後は音声認識のエラーによる評価の差と、実際にデジタルデバイドを実験参加者としたときの評価を行なう。

#### 参考文献

- [1] Ministry of Health, Labour and Welfare, “General condition of mental health caring and smoking countermeasure consequence (Japanese)”, 2008, <http://www.mhlw.go.jp/toukei/itiran/roudou/saigai/anzen/kenkou07/index.html>
- [2] R. Yaguchi., Y. Yoshida, “Project Success rate: 31.1% (in Japanese)” Nikkei Computer, pp. 36-53, 2008.
- [3] J. Weizenbaum, “ELIZA - A Computer Program for the Study of Natural Language Communication Between Man and Machine,” Communications of the ACM, vol.9, no.1, pp.36-45, 1966.
- [4] T. Shinozaki, Y. Yamamoto, S. Tsuruta, “Context-based counselor agent for software development ecosystem”, Springer Journal of Computing, 2013, doi: 10.1007/s00607-013-0352-y.
- [5] A. J. Gonzalez., S. Tsuruta., Y. Sakurai., J. Nguyen., K. Takada, and K. Uchida, “Using Contexts to Supervise a Collaborative Process,” Artificial Intelligence for Engineering Design, Analysis and manufacturing (AIEDAM), Cambridge University Press, pp. 1-16, 2010.
- [6] A.E. Ivey, N.G. Packard, M.B. Ivey “BASIC ATTENDING SKILLS 4th Edition”, 2006 by Microtraining Associates pp.93-97
- [7] K. Kurashige., S. Tsuruta., E. Sakurai., Y. Sakurai., R. Knauf., E. Damiani. “Context Respectful Counseling Agent integrated with Robot nodding for Dialog Promotion” IEEE International Conference on Systems, Man, and Cybernetics (SMC), 2017.
- [8] Fukushima O, Tsuchida T, Mori M, Matsumoto C, Suzuki A, “The effects of one-to-one interview method, group method, and assumed writing in counselor training program” Mejiro JPsychol3:63-75, 2007.
- [9] Likert R “A technique for the measurement of attitudes”, Arch Psychol 22(140):1-55, 1932.
- [10] Friedman, Milton. “The use of ranks to avoid the assumption of normality implicit in the analysis of variance”, Journal of the American Statistical Association. American Statistical Association. 32 (200): 675-701. doi:10.1080/01621459.1937.10503522. JSTOR 2279372, December 1937.
- [11] Friedman, Milton. “A correction: The use of ranks to avoid the assumption of normality implicit in the analysis of variance”, Journal of the American Statistical Association. American Statistical Association. 34 (205): 109. doi:10.1080/01621459.1939.10502372. JSTOR 2279169, March 1939.
- [12] Friedman, Milton, “A comparison of alternative tests of significance for the problem of m rankings”, The Annals of Mathematical Statistics. 11 (1): 86-92. doi:10.1214/aoms/1177731944. JSTOR 2235971, March 1940.
- [13] Wilcoxon, Frank, “Individual comparisons by ranking methods” (PDF). Biometrics Bulletin. 1 (6): 80-83. doi:10.2307/3001968, Dec 1945.