Designing an Intelligent Cognitive Assistant for Behavior Change in Mental Health

Tine Kolenik
Jožef Stefan Institute & Jožef Stefan International Postgraduate School
Jamova cesta 39
1000 Ljubljana, Slovenia
+386 1 477 3807
tine.kolenik@ijs.si

Martin Gjoreski
Jožef Stefan Institute & Jožef Stefan International Postgraduate School
Jamova cesta 39
1000 Ljubljana, Slovenia
+386 1 477 3812
martin.gjoreski@ijs.si

Matjaž Gams
Jožef Stefan Institute
Jamova cesta 39
1000 Ljubljana, Slovenia
+386 1 477 3644
matjaz.gams@ijs.si

ABSTRACT

The paper describes intelligent cognitive assistant technologies and how they can be used efficiently for behavior change in mental health, namely stress, anxiety and depression. It then reviews the state of the art of such cases, focusing on three different assistants and their experimental results. It raises some critical issues with the state of the art and the field itself, namely the lack of standardized evaluation and that current technologies do not take advantage of artificial intelligence and behavior sciences advances. The paper proposes its own comprehensive cognitive architecture for such an assistant, relying on the integration of behavior change theories and cohesive user modeling (together referred to as the theory of mind) into a cognitive architecture. In addition to linguistic input, the architecture includes biophysiological input for affect recognition purposes. Reinforcement learning and the use of the principle of multiple knowledge are proposed as the main drivers for strategy adaptation in relation to helping users. The paper attempts to fill the gaps in the works related to this field, which are mostly closed source, believing that an overview of the field, its issues and a proposed design will enrich the current academic landscape.

Keywords

Behavior change, cognitive architecture, intelligent cognitive assistant, mental health, user modeling and profiling.

1. INTRODUCTION

Intelligent cognitive assistants (ICAs) have been described as the next revolution in human-computer coexistence, particularly because of the idea of conversing with people in natural language [1]. The technology has relatively ancient roots in the history of artificial intelligence (AI), consisting of famous examples such as Weizenbaum's simulation of a Rogerian psychotherapist called ELIZA [2]. However, the technology has only recently laid the foundations for broad adoption in the form of ICAs such as Alexa and Siri as well as more domain-specific agents, and has been thus on the rise in terms of financing and research [1]. ICAs, which can be deployed as virtual agents or robots, are being made to: understand context; be adaptive and flexible; learn and develop; be autonomous; be communicative, collaborative and social; be interactive and personalized; be anticipatory and predictive; perceive; act; have internal goals and motivation; interpret; and reason. To achieve this in ICAs, they are embedded with a cognitive architecture (CogA), a "hypothesis about the fixed structures that provide a mind, whether in natural or artificial systems, and how they work together – in conjunction with knowledge and skills embodied within the architecture – to yield intelligent behavior in a diversity of complex environments" [3]. ICAs have recently shown great potential as persuasive technology in the domain of behavior change (BC). BC is a phenomenon that is considered to be a temporary or permanent effect on an individual in terms of their behavior, attitude and other mental states as compared to their past [4]. Persuasive technology can be defined as technology designed for attempting to "change attitudes or behaviors or both (without using coercion or deception)" [4, p. 20]. Persuasive technologies are already heavily used in the areas of health and wellness, where AI tracks people's behavior as well as physiological and mental states to motivate them to make better dietary decisions and exercise more along with offering people psychotherapeutic help in natural language [5]. To achieve this, user modelling and profiling (UMP) is extremely important in order to understand users' intentions, needs and states in relation to their psychographics, which can then be used for personalized and thus more effective BC through carefully selected outputs [6].

Of particular importance for persuasive technologies is the area of psychotherapeutic help. Stress, anxiety and depression (SAD) are prevalent and rising problems, with research revealing that more and more people are suffering from at least one of these mental issues, with figures in some groups reaching 71% for stress, 12 % for anxiety disorder and 48% for depression [7]. Slovenia, already struggling with the second highest suicide rate in Europe [8], also suffers from a serious shortage of officially recognized and certified mental health professionals as well as from a lack of regulation [9]. This opens the door to ICAs, as they are not only generally free to use (making help available to socioeconomically disadvantaged people) and available 24/7 (so people do not have to wait for their next therapy session), people tend to be more comfortable disclosing their feelings to an ICA than to a person [10], ICAs are available in remote locations, ICAs reduce burden on the healthcare system and its practitioners, and overall reduce barriers to mental healthcare access [11]. Such ICAs do not serve to replace professionals, but to complement them.

This introduction is followed by an overview of the state of the art. Afterwards, our design process and related work of constructing a psychotherapeutic ICA (PICA) are presented. Our work is in the phases of planning, analysis and design of the Systems development life cycle [12]. The paper focuses mostly on the computational (what the system does and why) and algorithmic level (how the system does what it does) of Marr's 3-level design hierarchy for AI systems [13]. Since papers on designs and their hurdles for PICAs are almost non-existent (as opposed to other kinds of ICAs), this work should find a useful place in the growing ICA research bibliography.

2. STATE-OF-THE-ART OVERVIEW

The chapter focuses on ICAs for BC in psychotherapy and mental health, namely for SAD. Various mechanisms of these systems are highlighted for the purposes of this work.

PICAs seem unique among ICAs, especially compared to Q&A ICAs. Users reveal personal information more freely, which makes PICAs more successful in their goals [14]. PICAs and their respective users also form a more longitudinal relationship. The interactions are not a one-off, where it is difficult to understand the users and act immediately with efficient strategies. This makes PICAs able to learn from historical interactions and improve in offering personalized psychotherapeutic help. However, since the use of ICAs as persuasive technology for BC is a recent endeavor, the pool of existing PICAs is small. The selection process of PICAs for this overview was based on the condition that they had been researched in an ecological environment (interacting with end users) with empirical experiments (e.g., randomized controlled trial). The metric is therefore being successful in relieving symptoms of SAD (a very indirect metric for determining which PICA performs best and why). PICAs cannot be compared in the same way as Q&A ICAs, where there are answers that need to be mapped to some collection of questions. PICAs must therefore be measured by their direct impact on the users [15], which makes them much more personalized in their acting, causing evaluation to be subjective and indirect. Our last selection metric was for PICAs to be text-based.

One of the most recent successful PICAs is Woebot [14]. Its overarching methodology is a "decision tree with suggested responses that accepts natural language inputs with discrete sections of natural language processing techniques embedded at specific points in the tree to determine routing to subsequent conversational nodes" [14, p. 3]. It gathers data on users' moods, goals, expectations and similar to build a user model and dispatch an intervention in the form of educational content, personalized messages, contextual strategies and scripted advice. In a randomized controlled trial, Woebot delivered better treatments to people than the government-approved self-help material, where the PICA relieved depression symptoms by app. 20% on average (Figure 1).

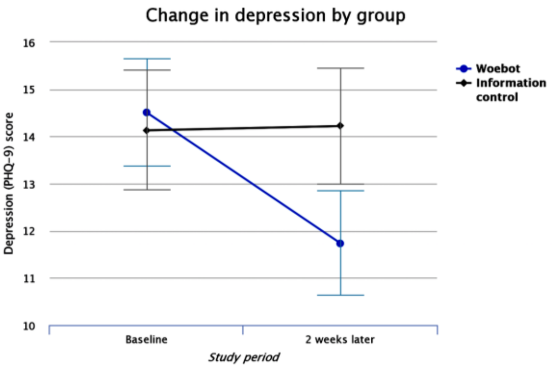

Figure 1 [14, p. 6]: Change in depression, control ('Information control') vs. test ('Woebot') group.

Another PICA is by Yorita, Egerton, Oakman, Chan and Kubota [16]. The PICA is built on the Belief-Desire-Intention (BDI) CogA (Figure 2). It comprises of three core models: "a conversation model for acquiring state information about the individual, measuring their stress level, a Sense of Coherence (SOC) model for evaluating the individuals state of stress, and Peer Support model, which uses the SOC to select a suitable peer support type and action it" [16, p. 3762], meaning it selects a suitable strategy based on the user model. 'Stress level' refers to scores from stress questionnaires, while 'state of stress' refers to three dimensions related stress that users actively try to improve: comprehensibility, manageability, meaningfulness.

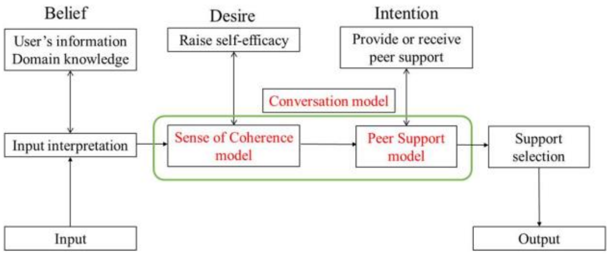

Figure 2 [16, p. 3762]: BDI CogA.

The experiment in the research found that the PICA had constructed more and more accurate models of its users, who managed stress better with each day of usage.

A PICA called Tess "reduce[s] self-identified symptoms of depression and anxiety" [17, p. 1]. It is crucially based on a wide emotion ontology, which it uses to determine and match users' emotions from their inputs. It uses stateful conversational models with scripted conversations, consisting of a natural language understanding module, dialogue state manager and natural language response generator. Improving strategies include asking for feedback as well as using user journaling data Tess gathers. Using Tess for psychotherapeutic interventions for mental health research showed that it significantly reduces anxiety and depression symptoms – roughly 16% for group 1 and 15% for group 2 (Figure 3).

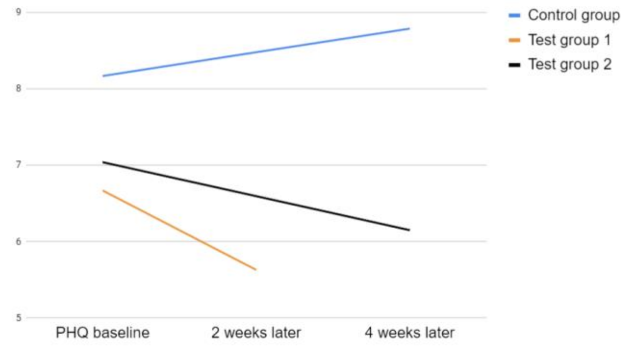

Figure 3 [17, p. 6]: Change in depression level (y-axis). Test group 1 used Tess for 2 weeks; 2 used it for 4 weeks. Control group used the government-suggested eBook on depression.

The PICAs described represent fully-formed agents with many mechanisms combined in a complete architecture. Most happen to be proprietary (closed source) and not available for technical scrutiny; there are no research papers that would describe them in detail, which holds true for most PICAs available online. Examining their architectures and implementation is therefore difficult, as noted by other researchers as well [18]. The overviewed work points at the PICAs being extremely goal-oriented and one-dimensional due to a lack of a contemporary UMP-based approach, a more coherent or theory-based CogA as well as a lack of inclusion if BC theories [5]. What follows is a description of a possible design and its hurdles towards a more effective PICA based on a coherent CogA, comprehensible UMP module and embedded BC domain knowledge.

3. THE PROPOSED DESIGN

This chapter describes the processes, related to the design of our PICA. First, an overview of contemporary methods of various fields used in ICAs is presented. Then, the necessity of embedded BC knowledge in relation to a comprehensive user model as well as a cohesive CogA for an effective PICA is argued through presenting the design and future implementation ideas.

For natural language understanding, dialogue-based applications are relevant for ICAs, where state-of-the-art applications use algorithms such as word2vec, Latent Semantic Analysis and deep learning methods [19]. Regarding UMP, there is a fairly accepted set of user characteristics that generally make artificial systems perform better when modelled: knowledge, beliefs, background, interests, preferences, goals, plans, tasks, needs, demographic information, emotional state, and context [20]. For CogAs, functional (as opposed to structural) architectures like BDI (and its various modifications) seem to be recommended [20].

Endowing a PICA with expert knowledge on BC has to be based on behavioral sciences advances in regards to human decision-making and similar phenomena [21], which has been recently combined with digital technologies, AI and big data. Many societal efforts are being put into creating persuasive technologies that would help, motivate, guide and affect people into bettering themselves and the world around them. One of the most powerful and effective contemporary BC concepts is the 'nudge theory'[1]. Nudge is "any aspect of the choice architecture that alters people's behavior in a predictable way without forbidding any options or significantly changing their economic incentive" [21, p. 6]. We believe that incorporating nudges and similar BC techniques into our PICA is essential for its effective behavior.

To effectively dispatch nudges and other strategies, our PICA (Figure 4) will build and continually update a user model. It is not clear which data on users is the most optimal for mental health BC, but we intend to implement a dialogically delivered questionnaire [22] on the Big Five personality traits (B5) as a fundamental element on the longitudinal (or global) user aspect [23]. This will largely inform PICA's BC strategies as B5 is an extremely successful psychological construct in determining what kind of influence is effective on a person as well as many other inferences [24]. In the less longitudinal aspects, determining SAD will be fundamental. This will again be achieved through dialogically delivered Depression Anxiety Stress Scales 21 questionnaire [25] to determine specific mental health symptoms. SAD scores will be refined on a regular basis, as reinforcement learning will be used to adapt PICA's behavior according to the changes in the SAD model [26]. The re-strategizing timeframe will have to be determined experimentally for optimal strategy adaptation. Continuous sentiment analysis (relying on Slovene sentiment lexicon JOB 1.0 if in Slovene [27]) will be used for a relatively short-term (albeit more inaccurate) inference on the user's emotional state [28]. Other user dimensions will comprise of as much of the previously listed accepted characteristics as is possible, feasible and sensible. Adding biophysiological measurements (e.g., heart rate, sweating rate and skin temperature) from smart bands is being considered as well. This may be used for automatic monitoring of SAD [29–32]. For example, at the beginning, the SAD answers can used to label the biophysiological measurements. Once enough data is labeled, the labeled biophysiological measurements can be used to learn personalized machine learning (ML) models for predicting SAD. The SAD predictions can be later utilized by the SAD model. This eases the user's burden since once the ML models for predicting the SAD are learned, the user will answer SAD questionnaires less frequently. The learning and the management of the personalized ML models will be handled by the Affect recognition module. Our PICA's strategy selection will be based on the principle of multiple knowledge [33] – which says that when high quality different viewpoints are sensibly filtered, adapted and combined, the result will be superior to the individual methods – and the predictive coding theory of cognition (PCT) [34]. There will be multiple competing strategies that will contend, adapt and enact according to multiple criteria from the interpreted linguistic input, user model and BC domain knowledge. A strategy (one response or a more long-term strategy) will be selected when a certain confidence threshold will be reached. If the threshold is not reached, necessary information on the user will be dialogically extracted until the probability that a certain strategy will be effective is reached. This approximately mimics PCT, but a strict formalization is needed in the future and further elaboration is currently out of scope for this paper.

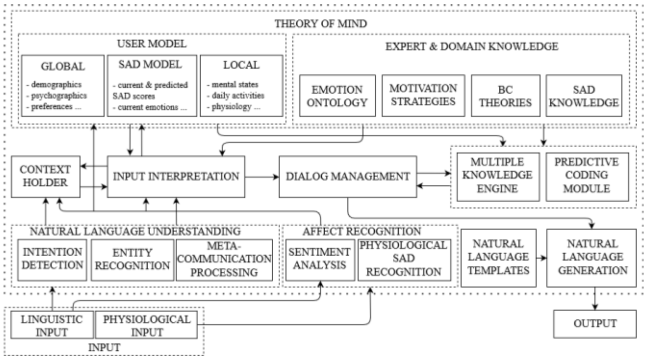

Figure 4: A tentative CogA for the described PICA.

4. CONCLUSIONS AND FUTURE WORK

This paper serves primarily the purpose of outlining the steps of planning, analysis and design of our PICA. This serves a variety of purposes: to highlight numerous problems in researching PICAs (the problems of existing PICAs being mostly proprietary and closed source, which means that there is a lack of research on PICAs that would illuminate technical details; a lack of standardized evaluation measures for PICAs), to contribute to a flourishing field of PICAs with a design paper, and to explicate our own unique ideas into a more coherent framework.

Our future work will require to define clear evaluation measures for PICAs. The lack of the latter points to a greater lack of a comprehensive overview on which CogAs are best suited for specific domains. This is all the more evident as there is little to no research on what user (or expert) data is needed for efficient PICAs. Our future work will continue with the Systems development life cycle steps (especially implementation; some work has already begun [35]) as well as with the third level of Marr's 3-level design hierarchy – physical (the realization of the first two levels) – after more thorough completion of the analysis and design steps pending feedback of the current work when published. The implementation will go hand in hand with experiments, which will inform the design of the CogA. The experiment by Fitzgerald et al [14] (see Woebot in chapter 2) will be replicated for comparison purposes.

[1] Richard Thaler, its author, received the Nobel Prize for it.

We believe that the highlighted research areas and technologies show promise and potential in both effectiveness (use case) as well as openness to new explorations and progress, which makes our research not only sensible, but also necessary.

5. ACKNOWLEDGMENTS

This work is supported by Slovenian Research Agency's Young researchers postgraduate research funding.

6. REFERENCES

[1] Oakley, J. 2018. *Intelligent Cognitive Assistants (ICA)*. Workshop Summary. IBM Almaden Research Center.

[2] Weizenbaum, J. 1966. ELIZA—a computer program for the study of natural language communication between man and machine. *Commun. ACM*. 9, 1 (1966), 36-45.

[3] Cognitive architecture. 2019. *Wikipedia*.

[4] Fogg, B. J. 2002. *Persuasive technology*. Morgan Kaufmann, Burlington, MA.

[5] Orji, R. and Moffatt, K. 2018. Persuasive technology for health and wellness: State-of-the-art and emerging trends. *Health Inform. J*. 24, 1 (2018), 66-91.

[6] Hardcastle, S. J. and Hagger, M. S. 2016. Psychographic Profiling for Effective Health Behavior Change Interventions. *Front. Psychol.* 6, 1988 (2016).

[7] Twenge, J. M. 2014. Time Period and Birth Cohort Differences in Depressive Symptoms in the U.S., 1982–2013. *Soc. Indic. Res*. 121, 2 (2014), 437-454.

[8] Just over 56 000 persons in the EU committed suicide. 2018. *Eurostat*.

[9] Kuralt, Š. 2015. Slovenija je po številu klinicnih psihologov na evropskem dnu. *Delo*.

[10] Lucas, G. M., Gratch, J., King, A., and Morency, L. P. 2014. It's only a computer: virtual humans increase willingness to disclose. *Comp. Human. Behav*. 37 (2014), 94-100.

[11] Mohr, D. C., Burns, M. N., Schueller, S. M., Clarke, G., and Klinkman, M. 2013. Behavioral Intervention Technologies: Evidence review and recommendations for future research in mental health. *Gen. Hosp*. 35, 4 (2013), 332-338.

[12] Sommerville, I. 2007. *Software Engineering* (9th ed.). Addison-Wesley Publishing Company, Boston, MA.

[13] Marr, D. 2010. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. The MIT Press, Cambridge, MA.

[14] Fitzpatrick, K. K., Darcy, A., and Vierhile, M. 2017. Delivering Cognitive Behavior Therapy to Young Adults With Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial. *JMIR Ment. Health*. 4, 2 (2017), e19.

[15] Provoost, S., Lau, H. M., Ruwaard, J., and Riper, H. 2017. Embodied Conversational Agents in Clinical Psychology: A Scoping Review. *J. Med. Internet. Res*. 19, 5 (2017), e151.

[16] Yorita, A., Egerton, S., Oakman, J., Chan, C., and Kubota, N. 2018. A Robot Assisted Stress Management Framework. *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Miyazaki, Japan, 2018*.

[17] Fulmer, R., Joerin, A., Gentile, B., Lakerink, L., and Rauws, M. 2018. Using Psychological Artificial Intelligence (Tess) to Relieve Symptoms of Depression and Anxiety. *JMIR Ment. Health*. 5, 4 (2018), e64.

[18] Stodden, V. 2011. Trust your science? *Amstat News*.

[19] Vinyals, O. and Le, Q. 2015. A neural conversational model. In *Proceedings of the ICML Deep Learning Workshop 2015* (Lille, France, July 10-11, 2015). IMLS, Lille, France.

[20] Sosnovsky, S. and Dicheva, D. 2010. Ontological technologies for user modeling. *Int. J. Metadata Semant. Ontol*. 5, 2 (2010), 32-71.

[21] Sunstein, C. and Thaler, R. 2008. *Nudge: Improving Decisions about Health, Wealth, and Happiness*. Yale University Press, New Haven, Connecticut.

[22] Rammstedt, B. and John, O. P. 2007 Measuring personality in one minute or less. *J. Res. Pers*. 41, 1 (2008), 203-212.

[23] IBM Personality models. 2019. *IBM Cloud*.

[24] Soldz, S. and Vaillant, G. E. 1999. The Big Five personality traits and the life course: A 45-year longitudinal study. *J. Res. Pers*. 33 (1999), 208-232.

[25] Lovibond, S. H. and Lovibond, P. F. 1995. *Manual for the Depression Anxiety Stress Scales (DASS)*. Psychology Foundation, Sydney.

[26] Fenza, G., Orciuoli, F., and Sampson, D. G. 2017. Building adaptive tutoring model using artificial neural networks and reinforcement learning. In *17th International Conference on Advanced Learning Technologies* (Timisoara, Romania, July 3-7, 2017). IEEE, 460-462.

[27] Bučar, J. 2017. *Slovene sentiment lexicon JOB 1.0, Slovenian language resource repository CLARIN.SI*. http://hdl.handle.net/11356/1112.

[28] Coppersmith, G., Dredze, M., and Harman, C. 2014. Quantifying mental health signals in Twitter. In *Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality* (Maryland, USA, June, 2014). ACL, 51-60.

[29] Gjoreski, M., Luštrek, M., Gams, M., and Gjoreski, H. 2017. Monitoring stress with a wrist device using context. *J. Biomed. Inform*. 73 (2017), 159-170.

[30] Wen, W. H. et al. 2018. Toward constructing a real-time social anxiety evaluation system: Exploring effective heart rate features. *IEEE T. Affect. Comput.* Early access (2018).

[31] Fedor, S., Ghandeharioun, A., Picard, R., and Ionescu, D. 2019. *U.S. Patent Application No. 16/168*, 378.

[32] Bolliger, L., Lukan, J., Luštrek, M., Bacquer, D. D. and Clays, E. 2019. Disentangling the Sources and Context of Daily Work Stress: Study Protocol of a Comprehensive Real-Time Modelling Study Using Portable Devices. *World Acad. Sci. Eng. Technol*. 13, 3 (2019), 3755.

[33] Gams, M. 2001. *Weak intelligence: through the principle and paradox of multiple knowledge*. Nova Science, Hauppauge, NY.

[34] Clark, A. 2013. Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behav. Brain Sci*. 36 (2013), 181-204.

[35] Mlakar, M., Tavčar, A., Grasselli, G., and Gams, M. *Asistent za stres*. http://poluks.ijs.si:12345/.